



1958

An MMPI Study of Religious Seminarians

Patrick John Rice

Loyola University Chicago

Recommended Citation

Rice, Patrick John, "An MMPI Study of Religious Seminarians " (1958). *Master's Theses*. Paper 1437.
http://ecommons.luc.edu/luc_theses/1437

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 1958 Patrick John Rice

AN MMPI STUDY OF RELIGIOUS SEMINARIANS

by

Patrick John Rice, S.J.

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

June

1958

LIFE

Patrick John Rice was born in Akron, Ohio, June 7, 1921. He was educated in Detroit, Michigan, at Gabriel Richard Grammar School and at Edwin Denby High School. He studied at the University of Chicago from 1938 to 1940 and at the University of Detroit during the year 1940-1941. On September 1, 1941, he entered the Society of Jesus at Milford Novitiate, Milford, Ohio. He pronounced perpetual vows as a Jesuit on September 8, 1943. In June, 1944, he was awarded the Litt. B. degree by Xavier University of Cincinnati, Ohio. In 1945 he began studies in philosophy and graduate studies in English language and literature in the West Baden College division of Loyola University. From 1948 to 1951, he taught Latin and English at Loyola Academy, Chicago, Illinois. In February, 1951, he was awarded the M.A. degree in English by Loyola University, with a thesis entitled "Maxwell Anderson's Theory of Tragedy." He also holds licentiate degrees in philosophy and in theology, awarded in 1948 and 1955 respectively by West Baden College. He studied theology at West Baden College from 1951 to 1955, and was ordained to the Roman Catholic priesthood there on June 14, 1954. In September, 1956, he began graduate studies in clinical psychology at Loyola University.

TABLE OF CONTENTS

Chapter	Page
I. STATEMENT OF THE PROBLEM.....	1
Traditional concern of Church about natural qualifications in candidates for priesthood--Directives of recent popes-- Growing interest in psychological techniques for screening candidates--The Bier MMPI study of seminarians--The hypotheses to be tested in this study.	
II. REVIEW OF RELATED LITERATURE.....	12
Rationale of the MMPI--Meaning of validity with regard to MMPI--Meaning of reliability with regard to MMPI--Construction of individual MMPI scales--Miscellaneous studies of seminari- ans--MMPI studies of seminarians.	
III. DESIGN OF THE RESEARCH.....	47
Subjects--Definition of terms "religious seminarians" and "adjusted seminarians"--Materials used--Procedure followed in the research.	
IV. ANALYSIS OF RESULTS.....	56
Homogeneity of the Experimental Group--Profile of the Ex- perimental Group--Comparison of the Experimental Group with Minnesota Male Normal Group--Comparison of the Experimental Group with the Bier Group of Seminarians.	
V. SUMMARY.....	73
REFERENCES.....	75

LIST OF TABLES

Table	Page
1. DISTRIBUTION OF EXPERIMENTAL GROUP (N=73) ACCORDING TO APPROXIMATE AGE AND YEARS OF TRAINING.....	57
2. DIFFERENCES BETWEEN MEANS OF NON-TEACHING GROUP (N=36) AND TEACHING GROUP (N=37) IN TERMS OF MMPI RAW SCORES WITHOUT K	58
3. NUMBER OF PROTOCOLS OF EXPERIMENTAL GROUP (N=73) WITH ABNOR- MAL ELEVATIONS (T>70) ON BASIS OF SCORES WITH K-CORRECTION ADDED.....	65
4. COMPARISON OF EXPERIMENTAL GROUP (N=73) WITH MINNESOTA NORMAL MALES ON BASIS OF RAW SCORES WITHOUT K.....	68
5. COMPARISON OF EXPERIMENTAL GROUP (N=73) AND BIER GROUP (N=171) ON BASIS OF T-SCORES TRANSFORMED FROM RAW SCORES WITHOUT K...	69

LIST OF FIGURES

Figure	Page
1. PROFILE OF EXPERIMENTAL GROUP (N=73) BASED UPON TRANSFORMATION OF RAW SCORES (x) WITHOUT ADDITION OF K.....	60
2. PROFILE OF EXPERIMENTAL GROUP (N=73) BASED UPON TRANSFORMATION OF RAW SCORES WITH ADDITION OF K (k).....	62

CHAPTER I

STATEMENT OF THE PROBLEM

The major religions of historical times have demanded of candidates for the priesthood a high degree of human perfection. A brief summary of prescriptions for aspirants to the priesthood of the ancient cults has been given by Guadier (24). The Roman Catholic Church has not shown itself less exacting in its demands upon those who would be official intermediaries between God and man. St. Paul, writing to Timothy and to Titus on the virtues required in a priest (I Tim., 3:2-7; Titus, 1:5-9) enumerated only human qualities -- evidently considering that these needed more elucidation than the requirements of supernatural motivation and supernatural virtues. Throughout the history of the Catholic Church, St. Paul's outline of requisite natural qualities has been filled in with specific details. Early in the Christian era, ecclesiastical councils and individual bishops concerned themselves with physical defects which were to be considered impediments to ordination. Later, epilepsy and other serious mental illnesses received consideration, as Geraud has pointed out (21). The latest revision of the Code of Canon Law forbids the ordination of those who "by reason of corporal defect cannot with security because of their weakness, or with decency because of difformity, exercise the ministry of the altars; also those who are or who have been epileptics or deprived of reason" (10, C. 984). But the Code of Canon Law determines only the lower limits of fitness for the Catholic priesthood.

Encyclical letters of the popes and decrees of ecclesiastical congregations insist that bishops should not be content with the mere absence of grave defects but ought to seek for positive signs of aptitude in candidates for the priesthood.

Among instructions on selection of candidates for the priesthood, preeminent are the encyclical of Pope Pius XI Ad catholici sacerdotii (44) and the encyclical of Pope Pius XII Menti nostrae (45). Pope Pius XII warns: "Achievement in the erection and management of Seminaries for the education of future priests deserves all praise. But it would be of little avail, were there any lack of care in the selecting and approving of candidates" (44, p. 28). He tells all who have charge of seminaries that "they must indeed foster and strengthen vocations with sedulous care; but with no less zeal they must discourage unsuitable candidates, and in good time send them away from a path not meant for them" (44, p. 28). He describes what he means by unsuitable candidates: "Such are all youths who show a lack of necessary fitness, and who are, therefore, unlikely to persevere in the priestly ministry both worthily and becomingly" (44, p. 28). Finally, he counsels severity rather than leniency in the matter of the selection of seminarians:

In these matters hesitation and delay is a serious mistake and may do serious harm. It is far better to dismiss an unfit student in the early stages; but if, for any reason, such dismissal has been delayed, the mistake should be corrected as soon as it is known. There should be no human consideration or false mercy. Such false mercy would be a real cruelty, not only towards the Church, to whom would be given an unfitted or unworthy minister, but also towards the youth himself; for, thus embarked upon a false course, he would find himself exposed to the risk of becoming a stumbling block to himself and to others with peril of eternal ruin (44, p. 28).

Later in the same encyclical, Pius XI cautions that the presence of a true

priestly vocation "is not established so much by some inner feeling or devout attraction, which may sometimes be absent or hardly perceptible; but rather by a right intention in the aspirant, together with a combination of physical, intellectual and moral qualities which make him fitted for such a state of life" (44, p. 29). He urges that superiors of seminaries and spiritual directors and confessors "reflect how weighty a responsibility they assume before God, before the Church, and before the youths themselves, if they do not take all means at their disposal to avoid a false step" (44, p. 29).

The reigning pontiff, Pius XII, in his exhortation to the clergy of the entire world on the sanctity of the priestly life, insists that the choice of candidates for the priesthood should be enlightened and prudent. He says: "it is always necessary to investigate individual aspirants to the priesthood with diligence, to ascertain the intentions and the reasons with which they have taken this resolution" (45, p. 34). Speaking specifically of the greater need for examination of candidates in recent years, Pius XII continues: "We deem it useful to exhort you to examine with your acknowledged prudence and with care whether those who wish to receive Orders are physically fit, all the more so because the recent war has not infrequently left deadly traces on the rising generation and has disturbed them in many ways. For this reason, these candidates should be carefully examined, and where necessary, the judgment of a good physician should be sought" (45, p. 35).

These quotations manifest the great concern of recent popes to improve the choice of candidates for the priesthood. Quite obviously, the pontiffs realize that the essential element of a vocation is not a natural quality or any combination of natural qualities, but rather the presence of the "actual grace"

of vocation -- a supernatural illumination of the mind, leading a man (but not forcing him) to think of the priesthood, and a supernatural strengthening of the will, leading him to desire the priesthood. Grace, however, is an internal spiritual, intangible entity, incapable of measurement. As such, it cannot be an objective criterion by which a seminary-director or a religious superior or a bishop can judge the presence of a vocation in another. Yet objective criteria of some sort are needed because a vocation needs official ecclesiastical approval. The Church demands that religious superiors or bishops or their delegates pass judgment upon the physical, intellectual, and moral qualifications of candidates for the priesthood.

The popes do not speak explicitly of such a quality as "psychic" fitness as a requisite in candidates for the priesthood. A fortiori, they do not speak explicitly of psychological examinations for prospective seminarians. In fact, Pius XI would seem, at first hearing, to have forbidden such examinations. In attacking the pervasive naturalism of the modern world, he has said: "But what is worse is the claim, not only vain but false, irreverent and dangerous, to submit to research, experiment and conclusions of a purely natural and profane order, those matters of education which belong to the supernatural order; as for example questions of priestly or religious vocation, and in general the secret workings of grace which indeed elevate the natural powers, but are infinitely superior to them, and may nowise be subjected to physical laws" (43, p. 25). It might be concluded from this text that the pope was forbidding such a procedure as subjecting candidates for the priesthood to psychological examination. To interpret the text so strictly would, however, be to call into question the procedure of the world-wide congress on religious

life held in Rome in 1950 (1). The delegates to that congress -- many of whom were persons close to the reigning pontiff, Pius XII, and well aware of his attitudes -- concerned themselves particularly with the question of whether modern sciences in general, and psychology in particular, could facilitate the selection of candidates. The proceedings of the congress were carried on with ecclesiastical knowledge, and detailed reports thereof have been published with ecclesiastical approbation. Moreover, as Benko and Nuttin argue (2), Pius XI himself often insisted upon the necessity of examinations of the intelligence and "temperament" of candidates for the priesthood. But such examinations are psychological examinations. Consequently, Pius XI was not opposed to psychological examinations as such. Benko and Nuttin further argue:

Pius XII draws attention to the possible necessity of consulting a physician, a layman Let us go one step further and ask this question: if one trusts the physician when there is question of an examination of the body, of physical health; if one judges that his advice constitutes in some fashion a criterion of the fitness of the candidate for the priesthood; for what reason would one not trust the examination of intelligence, of temperament, and -- to a certain extent -- of intention to a psychologist? (2, p. 5).

Pius XI seems to be saying merely that all the natural qualifications in the world do not constitute a vocation to the priesthood. A vocation is not directly discernible or measurable by scientific means. A young man may have the finest intellectual, physical, moral, and psychic balance and still not have a vocation to the priesthood. On the other hand, a vocation to the priesthood will not ordinarily be given by God to a man who does not have at least the minimal human qualities required for the proper fulfillment of the obligations of the priesthood. Consequently, if a young man is found to lack such minimal human qualities, it may prudently be concluded that he does not

have a priestly vocation.

Benko and Nuttin, throughout their recent publication on the subject (2), make a pointed distinction between the psychological examination of a vocation -- which is forbidden by Pope Pius XI -- and the psychological examination of a candidate for the priesthood. The examination of candidates for the priesthood has been required since the very beginning of the Church's history. In recent years, as scientific psychology has come of age and as psychic imbalance in the general population has attracted increased interest, churchmen have become increasingly alert to the possibility of psychic imbalance in candidates for the priesthood and increasingly receptive to the possibility of using psychological techniques in the examination of candidates. Benko and Nuttin (2) point out the particular need for what they felicitously call "psychological equilibrium" in the priest:

Certainly one might ask in what measure psychological equilibrium constitutes, for religious or priestly life, a condicio sine qua non. The practice of genuine virtue, and even great achievement therein, are possible in the man who is unbalanced and even in one who is deceived about his fundamental motives. Nevertheless, it is wise to note two things. Disequilibrium is accompanied by constructive qualities only in the "rich" personality, in whom psychic trouble consists less in a lack of maturity than in an unequal development and disproportion in certain traits. That is the essential distinction. Psychological equilibrium does not mean total absence of certain "excesses"; it is not a question of the state of soul which characterizes a type of man who is completely at rest. Psychological equilibrium is a question of degree of maturity and of integration. Furthermore, in the case of the priest, who fulfills in the midst of men one or other apostolic function -- the function of direction or education -- certain forms of psychological disequilibrium can be harmful, not only to the dignity of his function, but even more to the well-being of the persons towards whom he directs his apostolate. Even if one admit that the man who is psychologically unbalanced can attain, eventually, in the religious life a high degree of sanctity and personal perfection, the priest in contact with the world of men will not normally be able to shoulder the re-

sponsibilities of the apostolate unless a constellation of human qualities sustains his supernatural life. That is why it is important, it seems to us, to neglect nothing which could improve the methods of psychological diagnosis of the candidate for the priesthood. (2, pp. vii-viii).

Because many other ecclesiastics agree with the contentions of Benko and Nuttin, the practice is slowly gaining ground of screening candidates for the diocesan seminary and candidates for the novitiate of clerical religious orders with the aid of psychological tests. In the United States today, the practice is fairly widespread in the larger religious orders -- including the order of which the present writer is a member. The reigning pontiff is certainly informed of such practice. Moreover, results of studies of such screening have been published and have received considerable attention (2, 3). Projected screening programs involving the use of psychological examinations have been outlined in some detail in periodicals which are subject to ecclesiastical censorship and published with ecclesiastical approbation (4, 5, 11). A book written at the Pontifical Gregorian University in Rome on the subject of spiritual guidance and the psychological examination of "temperament" was published in 1956 with ecclesiastical approval (48). And, perhaps most significantly, the Acta of the worldwide congress on religious life held in Rome in 1950 contain a number of unchallenged pleas for the use of psychological tests in the selection of candidates for the priesthood.

The slowly growing interest of bishops, seminary staffs, and religious superiors in the application of psychological techniques is, then, quite unmistakable. Those who employ psychological tests as a help in selecting candidates are attempting to forecast those whose human qualities may render them more fit or less fit for the psychologically exacting life of the priest

in the modern world. In so doing, the proponents of such testing programs sincerely believe that they are fostering the spirit, if not following the letter, of the papal directive which urges that they "take all means at their disposal" (144, p. 29) to help provide the Church with suitable priests.

Hypotheses. One of the tests being used increasingly in the screening of candidates for the priesthood in the United States is the Minnesota Multiphasic Personality Inventory (MMPI), which was first published by S. R. Hathaway and J. C. McKinley and their colleagues of the University of Minnesota in 1943. The MMPI is "a psychometric instrument designed ultimately to provide, in a single test, scores on all the more important phases of personality" and to assist the investigator "to assay those traits that are commonly characteristic of disabling psychological abnormality" (27, p. 5). The rationale underlying the MMPI will be discussed at greater length in Chapter II of this thesis, which proposes to investigate the validity of the use of the test for the purposes of screening candidates for the priesthood within a religious order of the Roman Catholic Church.

In 1948, the Rev. W. C. Bier, S.J., published "A comparative study of a seminary group and four other groups on the MMPI" in Studies in Psychology and Psychiatry from the Catholic University of America. The study has since been republished as a chapter entitled "A comparative study of five Catholic college groups on the MMPI" in Basic Readings on the MMPI in Psychology and Medicine, edited by George Schlager Welsh and W. Grant Dahlstrom (50). Fr. Bier's group of seminarians (henceforth to be referred to as the Bier Group) was a rather heterogeneous sample drawn both from the ranks of diocesan seminarians and from the ranks of three different (unidentified) religious

orders in three geographically separated sections of the United States (3). All were "major" seminarians, i.e., men who had completed "minor" seminary studies at the high school level and were engaged in the study of philosophy or theology at the time of Bier's research. He compared this seminary group with four other Catholic populations (medical, dental, law, and undergraduate college students). Most of his subjects answered the MMPI anonymously. When he had rejected the protocols of those whose scores on one of the validating scales of the MMPI were in excess of two standard deviations above the Mean of the Minnesota Normal Male Group (the standardizing male group of the MMPI), Bier was left with an N of 171 seminarians. He found that 40% of this group showed scores on one or other clinical scale of the MMPI in excess of two standard deviations above the Mean of the Minnesota Normal Male Group. He concluded:

The seminary group manifests the same deviant tendencies [with respect to general male population norms for the MMPI] though in a more marked degree than the other groups [i.e., the Bier medical, dental, legal, and undergraduate groups] If the .05 level of significance is accepted, 55 per cent of the differences between the seminary and the other groups are significant; 40 per cent of such differences are significant at the .01 level. Of these statistically significant differences, 80 per cent are in the direction of greater deviation, i.e., poorer adjustment, for the seminary group. In other words, the seminary group is the most deviant portion of an already deviant population [i.e., the Catholic college and professional school population] (3, p. 593).

Bier's seminary group, then, differs significantly from both the Minnesota standardization sample and from a group of undergraduate and professional students in Catholic educational institutions. Bier expresses the belief that his seminarians constituted "a good representative sampling of students for the priesthood." (3, p. 588). This thesis proposes to investigate the

implications of that statement about the representativeness of the Bier seminary sample. It proposes, in other words, to ask Are Bier's results with seminarians on the MMPI so representative that they may be used without modification as norms for other seminary groups? Would all seminary groups perform as the Bier Group performed on the MMPI? Are other seminary groups as deviant from the general population norms for the MMPI as the Bier Group was? Might a seminary or religious order which proposed to use the MMPI as a screening device for candidates merely adopt the Bier norms in interpreting the MMPI performances of such candidates? Would candidates whose scores exceed the Bier norms be eo ipso suspect of unsuitability for seminary life or religious life?

The writer of this paper has tried to evaluate the validity of making such generalizations from Bier's data by applying the MMPI to a more homogeneous seminary population which should theoretically lie within Bier's "good representative sampling of students for the priesthood." He has chosen for his study a group of major seminarians from one religious order which has been using the MMPI as a screening device for the past several years and which is trying to refine its norms for scoring and interpreting the MMPI protocols of candidates. His subjects are drawn from a narrower geographical area than were the Bier seminarians. They are also fewer in number.

The investigator has set three null hypotheses to be investigated: (1) That there is no difference significant at the .05 level of confidence between the MMPI performance of the Bier heterogeneous seminary group and the more homogeneous group of this study; (2) That there are no intra-group differences

within this more homogeneous group significant at the .05 level of confidence; and (3) That there is no difference significant at the .05 level of confidence between the performance of this more homogeneous group and the standardization group of Minnesota Male Normals on the MMPI. More specific definition of some of the terms of these hypotheses and specific explanations of the methods employed in testing the hypotheses will be found in Chapter III of this thesis.

CHAPTER II

REVIEW OF RELATED LITERATURE

A. STUDIES ON THE MMPI

To say that the literature on the MMPI is extensive would be to risk understatement. Welsh and Dahlstrom (50) in 1956 published a bibliography of 689 articles on the MMPI, over half of which have appeared in the years since 1950. As they point out, "it is becoming increasingly difficult for a new student of the test to identify and locate the papers that will provide him with a proper foundation in the use of the MMPI" (50, p. v). For that reason, they have collected 66 articles which they believe constitute "the major research and clinical developments in the use of the MMPI during the last fifteen years" (50, p. v). Most of the MMPI literature reviewed in this thesis is contained in the Welsh and Dahlstrom compilation, although some of it was originally published in journals (to which Welsh and Dahlstrom give the references). In order to keep the review of literature within reasonable bounds, this investigator will restrict himself to reviewing only those studies which seem most basic: studies of the underlying rationale of the MMPI and studies concerned with the construction and with the validity of the individual clinical scales of the test.

Rationale of the MMPI. Paul E. Meehl, one of the men most closely connected with the development of the MMPI, admits that the MMPI may be classified

as a "structured personality test" (38, p. 5) and agrees with Max L. Hutt, who defines such tests as "those in which the test material consists of conventional, culturally crystallized questions to which the subject must respond in one of a very few fixed ways" (cf. 38, p. 5). Meehl, however, strongly maintains that there are "certain rather prevalent misconceptions as to the nature and the theory of at least one important structured personality test [the MMPI]" (38, p. 11) and sets out to correct such misapprehensions. He points out that, each of the test questions will have the same meaning to all subjects" (38, p. 5). He furthermore denies that the MMPI is a self-rating device which acts as a "surrogate for a behavior sample" (38, p. 6). He is at pains to point out, too, that the MMPI was not constructed upon an a priori basis. In short, Meehl insists that the MMPI represents a departure from what may be called the traditional approach to structured personality tests. Traditionally, a personality test of the structured (as opposed to the projective) type is constructed from a number of items assumed to be indicative of one or other personality trait. It is traditionally assumed that the items will mean the same thing to all subjects, and that certain items will be answered by certain types of persons in certain ways. Meehl says: "The fallacious character of this procedure has been sufficiently shown by the empirical results of the MMPI alone" (38, p. 6). He goes on to describe a second possible approach to the construction of structured personality tests -- the approach which was in fact taken by the authors of the MMPI:

The second approach to verbal self-ratings is rarer among test-makers. It consists simply in the explicit denial that we accept a self-rating as a feeble surrogate for a behavior sample, and

substitutes the assertion that a "self-rating" constitutes an intrinsically interesting and significant bit of verbal behavior, the non-test correlates of which must be discovered by empirical means. (38, p. 6; *italics mine*).

As examples of such an approach to testing, Meehl singles out the Strong Vocational Interest Blank, the Humm-Wadsworth Temperament Scales, and the MMPI -- "or any other structured personality measuring device in which the selection of items was done on a thoroughly empirical basis using carefully selected criterion groups" (38, p. 6). In such an approach to test construction, the test-maker does not a priori decide that an item of such-and-such content should tap such-and-such a personality dimension or that the answer to such-and-such an item by a particular subject should be taken at face value as the objectively true statement of a fact. Meehl gives an interesting example which highlights the distinctive approach of the MMPI:

One of the items on the MMPI scale for detecting psychopathic personality (Pd) is "My parents and family find more fault with me than they should." If we look upon this as a rating in which the fact indicated by an affirmative response is crucial, we immediately begin to wonder whether the testee can objectively evaluate how much other people's parents find fault with them, whether his own parents are warranted in finding as much fault with him as they do, whether this particular subject will interpret the phrase "finding fault" in the way we intend or in the way most normal persons interpret it, and so on. The present view is that this is simply an unprofitable way to examine a question-answer personality test item. To begin with, the empirical finding is that individuals whose past history and momentary clinical picture is that of a typical psychopathic personality tend to say "Yes" to this much more often than people in general do. Now in point of fact, they probably should say "No" because the parents of psychopaths are sorely tried and probably do not find fault with their incorrigible offspring any more than the latter deserve. . . . Again, "Much of the time I feel I have done something wrong or evil." Anyone who deals clinically with psychopaths comes to doubt seriously whether they could possibly interpret this item in the way the rest of us do, but they say that about themselves nonetheless. Numerous other examples . . . appear on the same scale and are significant because psychopaths tend to say certain things about themselves, rather than because we take these statements at face value (38, p. 7).

As Meehl says, in the case of the MMPI, "the stimulus situation seems to request a self-rating, whereas the scoring does not assume a valid self-rating to have been given" (38, p. 8). Meehl concludes what is perhaps the most basic of all articles on the MMPI with a perceptive analogy to clinical practice which suggests that the MMPI approach is a highly reasonable approach:

It has not been sufficiently recognized by critics of structured personality tests that what a man says about himself may be a highly significant fact about him even though we do not entertain with any confidence the hypothesis that what he says would agree with what complete knowledge of him would lead others to say of him. It is rather strange that this point is so often completely passed by, when clinical psychologists quickly learn to take just that attitude in a diagnostic or therapeutic interview. The complex defense mechanisms of projection, rationalization, reaction formation, etc., appear dynamically to the interviewer as soon as he begins to take what the client says as motivated by other needs than those of giving an accurate verbal report. There is no good a priori reason for denying the possibility of similar processes in the highly structured "interview" which is the question-answer personality test (38, p. 8).

W. Seeman has done an important study of "subtlety" in the MMPI (46). His work is primarily an investigation of items whose "meaning" in terms of diagnostic significance cannot be arbitrarily assigned in a priori fashion. He says:

As an example, consider two items from the MMPI: "It takes a lot of argument to convince most people of the truth" and "I have a habit of counting things that are not important, such as bulbs on electric signs and so forth." To the extent that the psychodynamic meaning of the second item can easily be established with a high degree of interpersonal agreement (that is, most individuals who have had the requisite psychological or psychiatric training would agree that this is an obsessive-compulsive mode of defense) whereas this is not true of the first item, the first item would by definition be properly characterized as "more subtle" than the second (46, p. 41).

Seeman remarks that "it is this property of subtlety in which structured personality instruments [like the MMPI] have been commonly presumed to be

deficient" (46, p. 42). He sets out to disprove the validity of this common presumption in the case of the MMPI by an experiment in which he used as subjects 58 advanced students in clinical psychology. He gave them 30 items from the MMPI, of which 15 were "obvious" items such as the counting item referred to earlier, and 15 were "subtle" in the sense that it was hypothesized that their "meaning" or psychodynamic significance would not ordinarily be detected even by diagnostically sophisticated persons. He found his prediction to be statistically verified in the sense that there was greater difficulty (significant at the .01 level) in assigning the "subtle" items to the correct diagnostic scale of the MMPI than in assigning the "obvious" items. If this was true for advanced students in psychology, it would presumably be even truer for the unsophisticated subjects to whom the MMPI is ordinarily administered. The experiment was repeated with the same items and the same subjects when, two academic quarters later, they had completed a course devoted to a study of the MMPI through readings and lectures given by Dr. Paul Meehl. Seeman again found that the degree of success achieved in the assignment of the obvious items was markedly greater than that achieved with the subtle items. He concluded that "even instruction in the MMPI by an individual thoroughly conversant with its structure does not close the gap between the subtle and the obvious items" (46, p. 49).

A study leading to somewhat similar conclusions about the "subtlety" of the MMPI was made by H. G. Gough (23), who also worked with sophisticated subjects. He conducted an examination of typical conceptions of neuroticism by having subjects take the role of a psychoneurotic patient in responding to the MMPI. He concluded: "significant discrepancies between what diagnosed

patients did, in fact, report on this inventory and the stereotypes given by simulators were discovered" (23, p. 57). He advanced the argument that "errors of the magnitude observed suggested a considerable degree of misinformation in the prevailing conceptions about neuroticism" (23, p. 57). Gough's study may be read as a confirmation of the contention that the MMPI is a subtle test, not as capable of being "faked" as it might at first sight appear to be.

S. R. Hathaway and J. C. McKinley, the authors of the MMPI, are obviously the most competent to speak on the subject of the construction of the test. They describe the formation of the original item-pool as follows:

The individual items were formulated partly on the basis of previous clinical experience. Mainly, however, the items were supplied from several psychiatric examination direction forms, from various textbooks of psychiatry, from certain of the directions for case taking in medicine and neurology, and from earlier published scales of personal and social attitudes. The original list consisted of more than one thousand items. . . . The separate items were formulated as declarative sentences in the first person singular. The majority were placed in the positive, the remainder in the negative. Interrogative sentences were not used. Simplified wording constituted the language of the items, the words used being selected as far as possible from those in most frequent use according to standard word frequency tables. Also, the statements were restricted to matters of "common knowledge." Idiomatic expressions were included when the idioms were common in the English language. Grammatical form was occasionally sacrificed in the interests of brevity, clarity, and simplicity (28, p. 60).

Items which seemed to be duplicates were deleted, as were items which seemed at the outset to have relatively little significance, so that the final number of items used in the first form of the MMPI was 504. For the sake of convenience, the items were arbitrarily classified under 25 headings, according to manifest content or "face validity." The process of construction of the test thus far described is, of course, an a priori process; before one can have a test, one must have items, and these items must come from somewhere. But the

a priori approach to the MMPI ends right there. Every other step in the construction of the scale was taken on the basis of empirical evidence.

Empirically, "a great many good basic pool items were eliminated from a final scale because they showed overlap in validity with some other clinical syndrome" (26, p. 104). Moreover, some items proved to have no empirically demonstrable clinical value; consequently, some items which remain in the MMPI are simply not scored in any way.

Once the basic item-pool had been assembled, the authors of the test went on to construct empirically various "scales" (sets of items which were empirically demonstrable as significant in connection with various clinical syndromes). As Hathaway says, "specific steps in scale development were so varied that they cannot be completely described" (26, p. 104). But McKinley and Hathaway do attempt to describe the purpose of the MMPI and the detailed process which, with minor modifications, was carried out in the formulation of each of the MMPI scales (35). Because the construction of the MMPI is considered to be a model of empirical procedure, and because Hathaway and McKinley speak with supreme authority on the construction of the MMPI, it will not be out of place to quote them exactly and at length:

The problems to be solved by the scales of the MMPI are frankly those of detecting and evaluating typical and commonly recognized forms of major psychological abnormality. The terminology and classification system are largely drawn from ordinary psychiatric practice. Where there are correlations between clinical syndromes, the scales tend to show correlation; where the clinically recognized diagnosis is impure, the scales will tend to be impure. These are usually, therefore, not statistically pure scales. . . . One additional point should be especially stressed. Every item finally chosen differentiates between criterion and normal groups and that is the reason for acceptance or rejection of the items. They are not selected for their content or theoretical import. Frequently the authors can see no possible rationale to an item in a given scale; it is nevertheless accepted

if it appears to differentiate. . . .

Specifically, the derivation of scales begins with the selection of a criterion group or groups. These persons have all been examined and diagnosed by the staff of the department of neuropsychiatry as patients in the inpatient service of the University [of Minnesota] Hospitals. The size of the criterion group varies usually between 25 and 50. For some scales it required several years to collect a sufficient number of cases to permit satisfactory scale derivation. These criterion cases are selected to be as representative as possible of the classical concept of the given syndrome. . . .

For each scale the responses of the criterion group or groups to each of the 550 items of the MMPI were tabulated to show the percentage frequency of occurrence of each possible answer -- True, False, Cannot Say. These response frequencies were tabulated for comparison with expected frequencies as determined on normal groups.

The normal groups most commonly used for item by item contrast were composed of 339 persons selected from among general Minnesota normals and of 265 precollege cases from among high school graduates applying for admission to the university. The general sample was divided into 139 men and 200 women, tabulated separately to show sex differences. These persons were between the ages of 26 and 43 inclusive and were all married. They declared themselves to be not under a doctor's care at the time of taking the inventory and are considered normal on that single basis. The modal number of years of schooling was 8 and few had gone beyond high school. These particular persons were used because they were felt most likely to be stable and representative. The tabulation for the entering college students was based upon 151 men and 114 women. These latter tabulations were invaluable in controlling the strong tendency of responses to certain items to vary widely in accordance with age or intelligence, or both.

For all scales the percentages for the criterion groups were compared with each of the normal percentages and an initial reservoir of items was selected which included all those showing a consistent difference. Statistically no item was chosen that showed a difference less than twice its standard error and most items yielded differences greater than three times their standard errors. . . .

To establish the validity of the various scales as they were derived, their power to differentiate test cases from normals was used as an indicator. "Test cases" is the term used in this paper to designate cases identified relatively or entirely independently of the criterion groups. For the most part, these cases were drawn from among hospitalized patients who were diagnosed routinely by the staff during the preliminary derivation of items and before any scale

was available. Where possible, test cases were taken from records and diagnoses made in an entirely different clinical setting. Naturally these latter cases are most desirable. . . .

It is important to note, nevertheless, that test cases were not so carefully selected as the criterion cases to represent either the pure syndromes or careful evaluation by the staff. . . . In considering the data presented showing the standard scores of test cases against the normal groups, it can usually be assumed that the data given represent a poorer picture than would be yielded if the cases could have been more carefully selected and the normals more adequately proved normal (35, pp. 87-89).

It will have become evident from the preceding quotation that validity of an MMPI scale is measured in terms of degree of discrimination between normals and abnormals of the type which a given scale is designed to identify. Since a T score of more than 70 is considered abnormal on an MMPI scale, validity is expressed in terms of the percentage of normals who score above 70 on a scale and the percentage of abnormals who score below 70 (or, in other words, in terms of the percentage of "overlap") rather than in terms of the more usual coefficient of validity. Where such data are available for a given scale, they will be noted in the discussion of the individual scales later on in this chapter.

If validity deals with percentage of overlap, what does reliability mean in connection with the MMPI? This is a question that is not easy for even the authors of the MMPI to answer. Most reliability studies of the MMPI have been test-retest studies. But, as McKinley and Hathaway point out: "It is pertinent to introject that the statistical thinking derived from aptitude and achievement testing should be amended when personality tests are considered. Many traits of personality are highly variable. Otherwise there would be little meaning to psychotherapy or preventive mental hygiene. Test-retest data

on MMPI scale are more a measure of trait variance than of reliability of scales" (35, p. 93). Consequently, when a coefficient of reliability is quoted in connection with a given MMPI scale, the careful student of the test will interpret such a coefficient in the light of what is known from other sources about the stability or variability of the trait which the scale is designed to measure.

Individual Clinical Scales. Scale 1 (Hs). The hypochondriasis scale of the MMPI is designed to detect "abnormal, psychoneurotic concern over bodily health" (34, p. 64). McKinley and Hathaway state that they hope to isolate by this scale those persons whose abnormality the American Psychiatric Association would classify under the heading "psychoneurosis, hypochondriasis." Thus, they say, they have "arbitrarily limited the statistical differentiation to the diagnostic group under the psychoneuroses and have excluded the symptomatic implications of the term as applied to the psychoses" (34, p. 64). Their criterion group contained, as far as they could determine, only pure, uncomplicated hypochondriacal cases -- to the number of 50. The normal group was composed of 109 males and 153 females (as described above, p. 20) and an additional 265 college students (entering freshmen). The Hs scale was originally a scale consisting of purely somatic items. This scale was considered insufficiently discriminating. Eventually a better scale (called the H - C_H Scale) was constructed; on this scale, 2% of the normals scored above T=70, and 40% of the diagnosed hypochondriacs scored below T=70. The scale was, therefore, about 60% efficient, with the possibility that 2% of those diagnosed psychoneurotically hypochondriac would be "false positives." With the development of Scale 3 (Hy) and of Meehl's K Scale, however, the

hypochondriasis scale was returned to its original somatic form. Hathaway explains the development of the scale into final form:

It had been noted in the course of derivation of a scale for hysteria that the new scale 3 (Hy) differed from the hypochondriasis scale mainly in the items related to the correction C_H for H. Scale 1 (Hs) was therefore arbitrarily made into a somatic item scale by eliminating the C_H items and some of the old H items that did not stand up on further analysis. This decision was intended to make the diagnosed hysterics score high mainly on scale 3. When K was tried on scale 1, the results showed that the corrected scale improved the differentiation between hypochondriasis and hysteria. It appeared that too extreme a purification had been made when all the C_H items were taken out. The addition of .5 K helped correct this error. . . . The C_H items correlated well with K; the correlation was negative because the items were scored inversely. In short, modern Hs + .5K is a compromise between a pure somatic scale and the old H - C_H . (34, p. 75).

Scale 2 (D). The D scale is intended to measure "symptomatic depression."

Hathaway and McKinley (29) say that the term "symptomatic" is used "because the authors wish to avoid the identification of the term 'depression' with anything other than the presence at the time of testing of a clinically recognizable, general frame of mind characterized by poor morale, lack of hope in the future, and dissatisfaction with the patient's own status generally" (29, p. 73). They point out that such a clinical picture might result from such divergent causes or occasions as economic or vocational frustration, personal problems, or the depressive phase of a cycloid mental illness. They note that such depression might represent a less stable trait in an individual than would, for instance, a measured hypochondriacal tendency. For that reason, the problem of obtaining their original criterion group was a difficult one. The group ultimately chosen consisted of 50 patients, "most of whom were in the depressed phase of a manic-depressive psychosis" (29, p. 74). They had all been "thoroughly investigated medically and psychiatrically and, as far as

possible, represented relatively pure cases of depression" (29, p. 74). From the protocols of these criterion patients, the D scale was developed in the manner previously described. Then a cross-validation study was made with 35 "test cases" (as defined earlier). It was discovered that "although there is some overlap, the scale yields scores that differentiate at least 50 per cent of the test cases from normals and even from other psychiatric cases, although the latter are, reasonably enough, more depressed than normal" (29, p. 78). Almost 7% of the normal group showed up as "false positives" on the D scale, however. Test-retest reliability would not be expected to be great for the D scale, since symptomatic depression can disappear and reappear quite quickly. For normals, however (40 cases reported by McKinley and Hathaway), test-retest study yielded a coefficient of $.77 \pm .044$, on the basis of which the probable error of a score is 1.9 points. The MMPI Manual uses the following phrases to describe the sort of personality traits the D scale is designed to identify: "poor morale of the emotional type with a feeling of uselessness an inability to assume a normal optimism with regard to the future"; "deep concern with the grim realities of life"; "lack of self-confidence, tendency to worry, narrowness of interests, and introversion" (27, p. 19). The Manual contends that "this scale, together with the Hs and Hy scales, will identify the greater proportion of those persons not under medical care who are commonly neurotic, as well as individuals so abnormal as to need psychiatric attention" (27, p. 19). The combination of Hs, D, and Hy scales has, in fact, come to be commonly known as the "neurotic triad" (50, p. 58).

Scale 3 (Hy). Almost at the outset of MMPI research, a promising preliminary scale for aid in diagnosing hysteria was developed (35). As time

went on, various attempts were made to improve the Hy scale. But, as the authors of the MMPI report, although the original scale was eventually somewhat bettered, most of the experimental hysteria scales were differentially less effective than the original "and it rapidly became apparent that [the] difficulty was due considerably to lack of definition in the clinical concept, to the concurrence of hysterical phenomena with other neurotic symptoms in the same individual, or to downright inability of the psychiatric staff to be sure of hysterical reactions in individuals who were under suspicion of developing organic disease" (35, pp. 89-90). The criterion group for scale 3 were persons who had received the diagnosis "psychoneurosis, hysteria" or who had been especially noted as having characteristic hysterical components in the personality disturbance which they manifested. McKinley and Hathaway point out again (35) in connection with scale 3 that, in the assignment of diagnostic terms, current clinical practice was followed as closely as possible. There was difficulty, however, in securing unanimity in diagnoses of hysteria: "Where cases showed a simple conversion symptom such as aphonia, an occupation cramp, or a neurologically irrational anesthetic area, the diagnosis was usually well agreed upon. In some cases there remained a doubt as to whether there was a true organic illness such as multiple sclerosis present or whether the syndrome reflected hypochondriasis or an early schizophrenic reaction" (35, p. 90). In the first stages of the construction of the Hy scale, the items that were found to be most discriminatory naturally grouped themselves into several categories: a large group referred to somatic complaints; another group consisted of statements tending to show that the patient considered himself exceptionally well socialized (35). The authors experimented

with the elimination of somatic items (to lower the intercorrelation between H_y and H_s), but found that such elimination "resulted in a marked drop in the number of test cases identified" (35, p. 90). Moreover, elimination of non-somatic items made the resulting H_y scale unduly sensitive to age and education. "These results forced the inclusion of some somatic items in the final scale, with considerable high correlation ($r = .52$ for normals and $r = .71$ for clinic cases) between H_s and H_y " (35, pp. 90-91). The authors point out that the H_y scale still retains some sensitivity to age and intelligence but that such sensitivity "seems valid clinically" (35, p. 91). They contend that the H_y scale discriminates the hypochondriac as an abnormal as well as does the H_s scale, but they point out that clinicians who have used both scales have seen a valid clinical difference between two persons having high scores on H_s and H_y but differing in that one score was higher: "There was a different prognosis and treatment indicated for the two. Where H_s was higher the physical complaints were diffuse and frequently required much less study to establish the presence of an important psychological factor in the disability. On the other hand, when H_y was dominant, the person frequently appeared normal psychologically and his physical complaints were likely to mimic closely or be accompanied by some common physical syndrome of the type now called psychosomatic" (35, p. 92). The raw score mean and standard deviation for 475 normal females were $M = 18.80$, $SD = 5.67$, and for 345 males they were $M = 16.50$, $SD = 5.50$ (35, p. 93). Test-retest data from 47 cases with an interval of three days to more than a year yielded an r of only .57. On a group of 98 high school girls retested after about one year the value was only $r = .47$. In explanation of these low coefficients of reliability, McKinley and Hathaway

say: "Although other objective tests have not proved the fact, clinically observed exacerbations and recessions of the symptomatic picture of hysteria in a given case are marked. An apparently normal person placed under sufficient strain will surprise everyone by developing symptoms. A case with a clear paralysis may get well momentarily and be undetected except on the basis of the history" (35, p. 93). Highest intercorrelations of the Hy scale are with the D scale (.55) and with the Hs scale (.52); lowest intercorrelation is with the Ma scale (.05). In summary, McKinley and Hathaway say: "This scale appears to measure a rather variable trait which is closely allied to and likely includes the earlier scale of hypochondriasis. The person who is especially characterized by Hy tends to be less obviously neurotic and to have, during disabled periods, a more specific set of physical symptoms" (35, p. 94).

Scale 4 (Pd). For the "psychopathic deviate" scale, the criterion group consisted of patients who were diagnosed "psychopathic personality, asocial and amoral type" (35, p. 98). They were of both sexes and ranged in age from 17 to 22 years. None was psychotic or neurotic, and most of the hysterical and clearly schizophrenic cases were eliminated. McKinley and Hathaway describe the criterion group in some detail:

The symptomatic backgrounds of the criterion cases were highly varied but can be characterized in several ways. Most often the complaint was stealing, lying, truancy, sexual promiscuity, alcoholic over-indulgence, forgery, and similar delinquencies. There were no major criminal types. Most of the behavior was poorly motivated and poorly concealed. All the criterion cases had long histories of minor delinquency. Although many of them came from broken homes or otherwise disturbed social backgrounds, there were many in whom such factors could not be seen as particularly present. Among the criterion cases there was a somewhat larger proportion of girls than of boys; this may have been due to the social selection that results from differential treatment by courts of boy and girl delinquents. (35, p. 98).

The authors report that "the scale was immediately valuable in the clinic," partly because of the "uncertainty of the average clinician when he attempts to examine a case of suspected psychopathic personality" (35, p. 99). Two groups of test cases were available for cross-validation: patients from the psychopathic unit of the University of Minnesota Hospitals and 100 male prisoners from a federal reformatory. The final scale is composed of items which fall naturally into several general categories: social maladjustment items (the most prominent group); items related to depression and the absence of strongly pleasant experiences; items suggesting paranoid trends. The authors remark that the items do not show a strong tendency to be highly intercorrelated; they conclude that "the final scale is, therefore, certainly not pure but deliberately mixed in factor content to yield greater clinical usefulness" (35, p. 100). For normals, the means and standard deviations of raw scores are $M = 13.44$, $SD = 4.23$ for 397 females, and $M = 12.99$, $SD = 4.00$ for 294 males. Only 5% of the Minnesota Normal Group scored higher than $T = 70$ on the Pd scale; of the federal reformatory test cases, 41% scored lower than $T = 70$ (so that the scale identified 50% accurately); of the clinic test cases, 55% scored lower than $T = 70$ (so that the scale identified 45% accurately). But it is to be remembered that the test cases were not so carefully selected as the criterion group, and that the scale was devised to identify only the asocial fraction of miscellaneous psychopaths. The authors note that the validity of the Pd scale is sharpened if the whole profile of each test case is considered. They say that "it is common for scores on other scales to be uniformly from one-quarter to one-half standard deviation distance below the mean, leaving the Pd score clearly dominant" (35, p. 101) in

the record of a test-case psychopath. They explain this profile phenomenon in the case of psychopaths in this way: "It is possible that this effect, which appears to be a general reduction in the measured abnormality, is produced by overly scrupulous, conscious avoidance of any betrayal by abnormal answers on the part of the subject. More likely these persons simply feel themselves to be overly perfect. Evidence for the latter suggestion lies in the fact that they seem clinically to be characterized by great self-esteem and self-interest (35, p. 101). McKinley and Hathaway cite a test-retest reliability coefficient of .71 obtained from a normal sample of 47 cases retested with an interval of a few days to more than a year (35, p. 102). Intercorrelations of the Pd scale with other scales for normals are highest with Sc (.60) and lowest with D (.29). The authors describe the cases identified by the Pd scale in the following detailed terms:

Most prominently the typical case has a shallow emotional life. The clinician may work very hard and become intensely interested in the patient but fail to receive in return more than a transitory and superficial loyalty. Sexual and other appetitive drives are not deeply effective in the patient's life. For example, although there may be promiscuity or actual prostitution, the female is frequently frigid and engages in sexual acts primarily as a means to social entertainment. Females are often masculine in interests. The psychopathic deviate seems to the observer to seek more and more dangerous or embarrassing experiences in the attempt to feel emotion like that of the normal. They sometimes commit suicide or more often nearly do so. This is again from shallow emotional sources rather than deep depression or normal recognition of failure. As they become older it is common for many of these cases to avoid more successfully real conflict with society. The lying, alcoholism, sexual promiscuity, or other behavior may persist; but it is somewhat more restrained and also society seems to feel less outraged. While these persons can usually verbalize as to the consequences of their behavior, there is often a failure to appreciate its significance for them in terms of their long-time social adjustment. Depression, when present, is usually expressed as fear of immediate punishment and loss of liberty rather than any reaction in guilt, regret, or the like. The tendency to blame others or to excuse themselves for their

predicament is common. They claim in self-extermination that they were misled by others who took advantage of their innocence, that the family discipline had been too severe so they rebelled, or some similar explanation. In clinical practice the Pd scale has been most valuable. So many of the cases with high scores are recidivists in delinquency that it is helpful to be put on guard. If the person is 16 to 19 years of age and has a score of twenty T points above most other scores on the profile, there is little likelihood that the person can stay out of trouble if not under rigid discipline. Older persons, however, more often avoid open breaks. In therapy, young persons with a high Pd should not be pushed toward maximal scholastic or vocational levels even when they have the capacities for training (35, pp. 102-3).

Scale 5 (Mf). The "masculinity-femininity" scale has turned out to be the least useful, because the least valid, of all the clinical scales of the MMPI. The "high 5" profile is the only type of profile not represented in the exhaustive Atlas (30) of profiles and corresponding case histories published by Hathaway and Meehl in 1951. Only two paragraphs on the construction of scale 5 appear in Welsh and Dahlstrom (50). Very few data are available on any aspect of the scale. The reasons for its general failure to perform significant diagnostic service are outlined briefly by Hathaway:

The difficulty in deriving a better Mf scale centered in the problem of a criterion by which the validity could be established. The published scale was derived by contrasting item frequencies from a small group of 13 homosexual invert males with those of average males as determined from the Terman and Miles I Scale with average males. A final less important criterion was the comparison of male and female frequencies. At first it seemed reasonable to collect relatively large samples of homosexual invert males and of homosexual females for more complete criterion evidence. The plan went awry because it became apparent that the homosexual samples were too heterogeneous. As we worked with the homosexual males and females, we came to feel that the groups were much more obviously divisible into several subtypes than was true for other clinical categories. For example, there is a pseudo-homosexual type where neurotic features related to inferiority seem to be dominant; there is a psychopathic variety with a strong tendency to high scores on Pd; and there is an invert group in which a constitutional factor seems probable. These and possibly other subgroups seem definite enough so that clinical study could separate them and much better

and purer Mf scales might be derived. Because the task was dependent upon having a comparatively large number of cases of each type and also because of the press of other research, this project was never finished (26, p. 110).

It is now rather widely thought that the Mf scale measures not so much femininity of interests but degree of education and culture. Evidence for such a viewpoint regarding scale 5 rests largely upon the fact that normal males who have been educated at the college level or beyond rather consistently produce scores in great excess of the scores of the Minnesota normal male population (3; 22). Such findings are understandable when it is recalled that the Minnesota normal males were largely rural residents whose modal educational history ended at eighth grade. Gough states that the MMPI was not found to be useful in studying cases of sexual deviation. He says: "Quite often profiles revealed high Mf scores with no discoverable clinical evidence of deviation, and the known homosexuals rarely attained a significant Mf score on the inventory. This raises again the possibility of homosexuality as divorced from a feminization of personality, and a feminine interest pattern with normality of sex behavior" (50, p. 345). Gough further reports that Harmon and Wiener found similar results with the Mf scale (50, p. 346). Although Hathaway wrote in 1956 that "the Mf scale has become widely used and . . . contributes considerably to routine clinical interpretation" (26, p. 110), he seems either unable or unwilling to say in what precise way it contributes. Hathaway's optimistic generalization seems clearly contradicted by the results of the specific studies mentioned above, and it is difficult to find in the literature any statement in praise of the Mf scale except Hathaway's own vague remark.

Scale 6 (Pa). The Pa scale was derived from criterion groups judged to be

showing paranoid symptoms. The most common diagnoses for the criterion subjects, according to Hathaway, were "paranoid state, paranoid condition, and paranoid schizophrenia" (26, p. 109). Symptomatically, the criterion groups "tended to have ideas of references (sic!), to feel that they were persecuted by individuals or groups, and to have grandiose self-concepts" (26, pp. 109-10). "Milder symptoms," says Hathaway, "included suspiciousness, an excess of interpersonal sensitivity, and an underlying rigidity of opinions and attitudes" (26, p. 110). Data on the validity and reliability of the Pa scale are sketchy. Hathaway admits that "cross-validation was always disappointing and the published scale was considered weak although it was the best that could be developed" (26, p. 110). Meehl's K correction, which is thought to have improved the diagnostic acuity of other scales, failed to sharpen the Pa scale. "It was felt," Hathaway explains, "that the K correction did not help because more than 20 per cent of the scale 6 items were already subtle in character" (26, p. 110). In certain instances, however, the Pa scale does have a restricted type of validity: "One factor that seemed to justify at least temporary use of the scale was that there were few false positives. When a person had a high score, he tended to be diagnosed as paranoid or at least he was felt to be sensitive and rigid in personal relationships" (26, p. 110). Apparently many clinicians feel that such validity makes the scale useful, since its "temporary use" seems to have developed into permanent use despite the fact that the scale remains unmodified.

Scale 7 (Pt). Scale 7 was constructed to identify patients typified by what used to be known as "psychasthenia." The general procedure for the construction of the scale differed only in minor details from the procedure

described earlier in this thesis. But McKinley and Hathaway state that this particular scale caused them considerable difficulty:

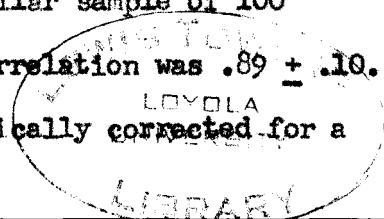
Unfortunately for the present study not many entirely satisfactory criterion cases of psychasthenia come into the closed wards of a psychiatric clinic. Many more are seen in the outpatient clinic or are advised by lay counselors and are never severely handicapped. Because we have felt unsure in the use of even carefully studied inpatients for purposes of scale derivation, we have avoided using criterion cases from the outpatient clinic. The criterion group is thus small and not entirely homogeneous. At least one of the cases appears to have been incorrectly diagnosed. Fortunately, the trait itself is the most homogeneous one so far described, so that correlations of items with the total score could be used as a guide (36, p. 82).

The criterion group consisted of 20 patients who had been intensively studied medically and psychiatrically and for whom the final diagnosis was psychasthenia in one form or other (36, p. 82). Psychasthenia is described by McKinley and Hathaway in the following terms:

The psychiatric classification of psychasthenia is applied to a group of individuals whose thinking is characterized by excessive doubt, by compulsions, obsessions, and unreasonable fears; these persons are often seen in psychiatric hospitals but are encountered much more frequently among normal groups by counselors and personnel workers. . . . Often a psychasthenic individual is characterized not so much by well-marked fears of individual things or acts as by great doubts as to the meaning of his reactions in what seems to be a hostile environment. In other cases the phobia becomes attached to certain acts or thoughts of the subject in such a way that he is forced through fear to compulsively perform needless, disturbing, or personally destructive acts or to dwell obsessively upon lines of thought which have no significance for his normal activities. Compulsive acts are always characterized by the need felt by the subject to perform them without regard to rational considerations. For example, he may always be forced to count objects or to touch a certain spot on a wall or to avoid stepping on sidewalk cracks. If he fails to do these things he feels uncomfortable; if he does them he is forced to rationalize and justify his acts. Obsessive thinking is itself commonly accompanied by anxiety so that the patient may be tense and anxious over the content of his thoughts as when he thinks over and over again that he is useless. Similarly, he may find himself anxiously obsessed with such ideas as the impending likelihood that he will faint or that something terrible or threatening is about

to happen. Again, he may be forced to think things which, while not in themselves producing anxiety, through his impatience and preoccupation with the fact that he cannot stop thinking them, do secondarily produce an anxious reaction The general reaction type characterized by these compulsive and obsessive acts and thoughts is called psychasthenia. The word derives from the concept of a weakened will that cannot resist the behavior regardless of its maladaptive character (36, p. 81).

Research with the scale disclosed "relatively little change in score with age" (36, p. 85). Some difference between the sexes was observed, but McKinley and Hathaway believe that "without further study no special significance should be attached to this difference," which is in the direction of higher scores for females of both the general normal and precollege samples (36, p. 85). For this scale, few test cases were available at the time that the research on construction of the scale was published; however, Hathaway and McKinley state that "nevertheless additional individuals so far obtained [as test cases] by clinical diagnosis have been deviates on the scale" (36, p. 85). The scale identifies 60% of the criterion psychasthenics when a T-score of 70 is used as a cutting-point; it identifies 95% of the criterion group when a T-score of 50 (the mean for normals) is used as a cutting-point. Therefore, "the evidence of validity as given by psychiatric cases with clinical symptoms of some degree of psychasthenia is relatively clear and positive" (36, p. 85). A test-retest measure of reliability yielded a coefficient of $.74 \pm .15$; the authors feel that "the coefficient obtained represents a low limit rather than a true test-retest correlation value" (36, p. 86). A split-half study yielded a coefficient of $.84 \pm .07$ for 200 random normal cases. When a similar sample of 100 psychiatric cases selected at random was used, the correlation was $.89 \pm .10$. When the two last-mentioned coefficients were statistically corrected for a



full-length test, they became $.91 \pm .07$ and $.94 \pm .10$ (36, p. 86). The Pt scale, therefore, may be said to have a high degree of reliability (which may mean, since split-half correlation was employed, something more than that the trait called psychasthenia is a stable trait). Pt is intercorrelated with the D scale more highly than with other scales, the correlation coefficient being $.44 \pm .10$ for normals and $.69 \pm .10$ for miscellaneous psychiatric cases. Hathaway and McKinley summarize the characteristics of the Pt scale by stating that it is "internally homogeneous" and that "further evidence of validity is given by the fact that, on the average, persons exhibiting psychasthenic symptoms to only a minor degree score significantly higher than normals" (36, p. 86).

Scale 8 (Sc). For the schizophrenia scale, two partly overlapping criterion groups were made up of patients who had been diagnosed schizophrenic. Hathaway says: "These cases were of assorted diagnostic subtypes and included about 60 per cent females and 40 per cent males. The final Sc scale was derived from a stock group of 152 items all of which showed statistically reliable differences for the schizophrenia criterion cases but many of which also differentiated depression cases, hypochondria cases, and other special groups" (26, p. 108). The Sc scale proved to be the hardest of all to sharpen and purify. Hathaway states: "From the very first it was found that differential cuts on cross-validation groups could not be pushed above a positive 50 to 60 per cent of the diagnosed cases identifiable with an apparently false positive rate of 10 or 15 per cent out of general normal cases" (26, p. 108). Therefore, an attempt was made to construct scales refined according to the subclassifications of schizophrenia: catatonic,

paranoid, simple, and hebephrenic. No significant improvement was achieved. However, says Hathaway: "The K scale, an outgrowth of Meehl's work on the derivation of a scale for a normal control component in behavior, finally provided a device by which the discrimination of the Sc scale could be sharpened. . . . The K correction raised the percentage of cross-validation cases reaching or exceeding T-score 70 to 59 and the corresponding percentage of normals dropped to 2 per cent Even with the correction, a considerable number of the cross-validation cases managed to stay below the T-score 61" (26, p. 109). Hathaway attempts a clinical explanation for the comparatively poor discriminatory power of the Sc scale:

Various investigators have found the clinical diagnosis of schizophrenia to be reproduced independently on the same patients by different clinicians in only 30 to 60 per cent of the cases. These figures are certainly not too low if less psychotic patients are used for the experiment. The MMPI Sc scale suggests, as do similar scales on other inventories, about the same degree of reproduction of the diagnosis in such clinical groups. There is no accepted way to assert that either the scale or the diagnosis is wrong. In the long run the decision should rest upon the useful correlates of test and diagnosis (26, p. 109).

The Sc scale has been reported to show high intercorrelation with scale 7 (Pt). Hathaway and Monachesi report a correlation of .68 between Sc and Pt; Cottle reports a correlation of .84 (26, p. 109).

Scale 2 (Ma). The "hypomania" scale attempts to isolate cases demonstrating "milder degrees of manic excitement occurring typically in the manic-depressive psychoses" (35, p. 94). According to McKinley and Hathaway, "the cardinal symptoms of maniacal conditions are generally stated to be an elated but unstable mood, psychomotor excitement, and flight of ideas" (35, p. 94). Hypomania -- the milder degree of mania which scale 9 attempts to identify --

follows a similar pattern but in lesser degree which "may be at times so unobtrusive as not to impress even an expert" (35, p. 94). Hypomania may be found among otherwise normal individuals in forms which McKinley and Hathaway describe at some length. Since elevations on scale 9 (Ma) are the most common elevations found among normals, it may be well to quote their description of the "high 9" normal individual:

Among normal individuals one may recall acquaintances who tend at times to be overtalkative, distractible, restless. Such a person may feel and appear to be extraordinarily well, enthusiastic, and energetic, but the use of his energy is likely to be inefficient because he tries to do too many things at a time. He is usually full of ideas which may be basically sound but they are not adequately worked out and if put into execution are seldom carried through to a satisfactory conclusion. Emotionally he may be a bit elated and too happy, he may be impatient and irascible or he may express ideas of feeling gloomy and somewhat frustrated; commonly the mood swings rapidly within minutes or hours from one to another of these attitudes, often without any corresponding environmental explanation for the shifts. Viewed over a longer period of time, it is often discernible that these persons tend to have periods of definite depression rather than elation or euphoria. Along with these characteristics, there is often egocentricity, lack of appreciation of the ineptitude of his behavior in given settings, and a certain obvious disregard for others (35, p. 94).

The criterion group for this scale consisted of 24 cases. McKinley and Hathaway state that "only manic patients of moderate or light degree were usable, since the more severe cases could not cooperate adequately" (35, p. 94). The clinical diagnoses were either "hypomania" or "mild acute mania," depending on the severity of the case. Care was exercised to exclude from the criterion group individuals with delirium, confusional states, or with excitements associated with other psychoses such as schizophrenia. Cases of agitated depression were likewise excluded. For the normal sample, the raw score mean and standard deviation from 379 females were $M = 13.65$, $SD = 4.50$; from 294

males, $M = 14.51$, $SD = 4.42$. The authors noted that among 900 available clinic cases, 30 received scores of 70 or more (i.e., T-scores) without any clinical note especially indicating hypomania. They maintain that "these cases also illustrate the tendency for psychopathic personality to be indicated by the hypomanic scale since 10 of them received this diagnosis or were chronic alcoholic cases" (35, p. 96). They also note "a tendency for cases with organic deterioration of the brain to receive high scores" on the Ma scale (35, p. 96). The authors conclude: "The evidence for the validity of the Ma scale is certainly not conclusive. There is, however, a tendency for persons with hypomanic symptoms to secure high scores. It is to be hoped that the scale would appear distinctly better if the criterion cases were better. This is one of several scales that will need to be checked further before final acceptance" (35, p. 96). McKinley and Hathaway report a slight negative correlation with the D scale among normals, as might be expected ($r = -.02$), and a degree of positive intercorrelation between scales Ma and Pd (.49) and between Ma and Sc (.56) for normals. They report a test-retest coefficient for Ma of .83 for normal subjects, from which they infer that "the trait has a surprising degree of stability in normal persons" (35, p. 97). The authors suspect that there may be two factors involved in Ma -- the one comparatively stable, the other transient: "The constant factor is likely to be something akin to what is commonly called optimism. Among our acquaintances, those whom we think of as optimists are rather consistently so, as are the pessimists. Apart from optimism there is also a variable tendency related to the usually episodic excitement of mania or hypomania which is seen in abnormal degree. The abnormal factor comes and goes and seems not to be strong among

normal persons" (35, p. 97). The authors note that the Ma scale has been useful in identifying "the juvenile delinquent, the overactive adult, and the agitated depression with ambivalent affect" (35, p. 98). They feel that "the delinquent with a high Ma score and lowered Pd has seemed more likely to benefit by counseling and by being given another chance," and they suggest that "the rather good prognostic indications in the adult case with an isolated Ma score are apparently in accord with general psychiatric opinion" (35, p. 98).

Scale 0 (Si). The tenth MMPI scale, numbered "0" and lettered "Si," aims to measure "the tendency to withdraw from social contact with others" (27, p. 21). The Si scale is not a clinical scale in the strict sense of the word, since it was neither standardized upon a criterion group of clinical cases nor intended primarily for use with persons suspected of clinical abnormality. The construction of the Si scale is the work of L. E. Drake (15; 16), who made an item analysis of the MMPI, using as an external criterion scores on the Minnesota T-S-E Inventory as scored for introversion-extroversion. On the basis of that criterion, students at the University of Wisconsin were divided into those who had obtained centile ranks of 65 and above (on the T-S-E) and those who had obtained centile ranks below 35. The N in each group was 50. The students were all females, although the scale was later validated with a male population. Items for the Si scale were selected because they showed a difference between the percentage responses of the upper and lower T-S-E groups of at least twice the standard error of the difference (15, p. 181). Twenty-eight of the 70 items on the Si scale appear on no other MMPI scales. Correlation of Si scores with T-S-E scores was $-.72$ for females and $-.71$ for males; the coefficient was negative because the key for the MMPI was

constructed so that a high score would indicate introversion whereas on the T-S-E a low score indicates introversion. The Si scale was later validated by comparing Si scores with number of activities engaged in by college students, who were also classified according to the sizes of the communities in which they had lived most of their lives (16, p. 184). Regardless of population, the means for students reporting two or fewer extracurricular activities were significantly higher (at the .0005 level of confidence) than the means for those who participated in four or more such activities. The scale is, therefore, considered to be a highly valid measure of introversion-extroversion as measured against two rather significant external criteria. Reliability coefficients are not reported.

Validating Scales. Besides the ten clinical scales enumerated and discussed above, the MMPI can be scored for validity on four scales, F, L, K, and ?. Since none of these validating scales enters significantly into this study, these scales may be dismissed rather briefly. The F scale ferrets out combinations of symptoms unlikely to exist in any one individual at any one time. "If the F score is high, the other scales are likely to be invalid either because the subject was careless or unable to comprehend the items, or because extensive scoring or rescoreing errors were made. A low F score is a reliable indication that the subject's responses were rational and relatively pertinent" (37, p. 18). The L scale consists of 15 items to which the vast majority of persons would tend to answer "Yes." "The L score . . . affords a measure of the degree to which the subject may be attempting to falsify his scores by always choosing the response that places him in the most acceptable light socially. A high L score does not entirely invalidate the other scores

but indicates that the true values are probably higher than those actually obtained" (27, p. 18). The K scale is thought to be a "measure of test-taking attitude" (27, p. 18). The Manual maintains that "a high K score represents defensiveness against psychological weakness, and may indicate a defensiveness that verges upon deliberate distortion in the direction of making a more 'normal' appearance" (27, p. 18). The K scale is considered to be more subtle than either the L or the F scale and to tap attitudes at a less conscious level which may serve to distort the scores (39). "A low K score tends to indicate that a person is, if anything, overly candid and open to self-criticism and the admission of symptoms even though they may be minimal in strength. A low K score can also result from a deliberate attempt to obtain bad scores or to make a bad impression" (27, p. 18). The value of the K scale is highly disputed. Meehl and Hathaway, of course, insist upon its value (39); they are seconded by Sweetland and Quay (50, p. 638). More or less opposed to the K scale are Gough (50, pp. 321-327), Schmidt (50, p. 636), de Beuchley & Ball (50, p. 622), and Hunt et al. (50, p. 630). The question score, or ? score, is determined simply by counting the number of items to which the subject fails to respond; it is felt that if 50 or more items are left unanswered, the record is invalid.

B. STUDIES OF SEMINARIANS

Miscellaneous Works. Most of the published material concerning the application of psychological techniques in the examination of candidates to the priesthood is foreign (particular French) in origin, theoretical in nature, and only tangentially related to the subject of this thesis. As Benko and Nuttin write, "the majority of these works are content with drawing the attention of

superiors to the importance of the problem and of giving them some practical suggestions" (2, p. 15). Benko and Nutton are referring to such works (listed among the references at the end of this thesis) as those of Biot (6), Galimard (7), Cahen-Salabelle (9), Cossa (11), Couly (12), Eck & Larere (18), Ernst (20), Geraud (21), Nabais (12), and Sinety (19). They single out the study of Sinety as "the first attempt to put at the service of superiors certain data from psychology" (2, p. 15). Cahen-Salabelle (9), a Jungian, feels that the psychological aspects of candidates, which he calls their affectivite, can contribute to their vocational adjustment or detract from it; but he does not believe that scientific psychological techniques can be of significant aid in detecting the role of affectivite among so specialized a group as seminarians. Cossa (11) describes the cases of four priests who, under psychotherapy, discovered that the motivation which led them into the priesthood involved considerable unconscious self-deceit; he maintains that two of them, after therapy, freely and consciously accepted their vocations on the basis of suitable motivation -- but he does not make clear what happened to the other two, except that none of the four has left the Church. Ernst (20), employing questionnaires and interview techniques, discovered that 40 per cent of the 30 subjects whom he examined had not placed a truly moral act in their decision to enter religious life; he concluded that the dynamism of unconscious motivation among seminarians needs considerably more study. Biot & Galimard (7) and Geraud (21) present convincing arguments for the opinion that sexual disequilibrium in a candidate for the priesthood constitutes an absolute counter-indication. These authors have been challenged only by Marc Oraison, whose book Vie chretienne et probleme de la sexualite (Paris, 1951) was

ecclesiastically condemned shortly after its publication (and, consequently, is not included among the references in this thesis). Couly (12) makes considerable point of the need for professional advisers for directors of seminaries because "it is the manias, the singularities, the minor traits which have no meaning to most men which are, on the contrary, for an experienced doctor, the certain index of an extremely dangerous mental illness" (12, p. 18). The unpublished study of Nabais (42) is concerned largely with the examination of the intelligence of prospective seminarians; he proposes two striking cases (from among 73 minor seminarians) in which psychological tests proved that the general impression of seminary directors and professors were gravely erroneous.

Among works published in the United States, perhaps none has been more influential than that of Duffey (17), who insists that the "spirit" of candidates for priestly and religious life must be "tested" -- but who is rather suspicious of attempts to test that spirit with psychological instruments. Duffey has been rather severely criticized by Bier (4) who contends (and demonstrates) that there is nothing in the theology of vocation which should prohibit the psychological testing of candidates. Bier (5) also criticizes the overly facile approach of Sr. M. Digna (14), who would be content with merely training a psychometrist for every religious congregation. The need for screening of candidates for the priesthood was perhaps most vividly presented by Moore (40), who, in 1936, published the results of an extensive scientific survey on the subject of insanity in priests and religious; Moore discovered that, if parietic cases were ruled out (because paresis, for obvious reasons, is an extremely rare condition among priests and religious), the rate of incidence of mental illness among priests and religious was greater than

among the general population. Perhaps because he suspects that mental illness among priests and religious could spring from inept spiritual direction, Simoneaux (48) has recently published an interesting attempt to correlate spiritual guidance with "varieties of character" (following the characterology of LaSenne and Heymans). Burke (8), studying minor seminarians by means of tests and rating scales (filled out by superiors), discovered that the most certain indices of probable success in the minor seminary are high results on achievement tests taken before entrance, high results in Latin and English at the end of the first year in the seminary, and favorable ratings by superiors at the end of the first year. McCarthy (33), using the inventories of Bell, Bernreuter, Allport-Vernon and others, discovered that seminarians in general show a more accentuated neurotic tendency than do lay students. In a well designed research study employing the MMPI, a sentence-completion test, and the Draw-a-Person test, Mother M. Elaine, M.C.S.A., has recently found (19) that religious women scored significantly less favorably than did four other related groups. Finally, in an unpublished study of religious women, the Rev. Richard P. Vaughan, S.J., has concluded that members of contemplative religious orders of women become apparently (when judged by general population norms) less well adjusted psychologically in proportion to the amount of time spent in religious life.

It will be obvious that most of the studies cited above are related to the subject of this thesis only in the sense that they indicate increasing interest in the psychological aspects of priestly and religious vocations. Only the two studies to be described below really parallel the present study in great detail.

MMPI Studies of Seminarians. The work of Bier (3), which has already been described in some detail as the point of departure for this thesis, contains an important aspect which has not thus far been discussed. Bier concluded that a number of items on the MMPI were either inapplicable or positively unsuitable to his seminary group. Among such items were some of a religious nature and some of a social nature (including items asking about activities, for example, which are forbidden or inaccessible to seminarians). As a result of a detailed item analysis, Bier made a rather significant statement which has since proved influential among those who wish to apply the MMPI to candidates for the priesthood:

It is suggested here . . . that some modification should also be introduced in the content of the MMPI [i.e., as well as in the statistical norms] in adapting it to seminary use. More specifically, it is suggested that certain items should be eliminated. This proposal is based upon the assumption . . . that certain MMPI items have no application to the seminary group and upon the experimental fact that a number of these items do not discriminate between the well-adjusted and the poorly adjusted seminarians. When these two criteria agree in picking out the same items, the author believes that such items can be eliminated from the test without loss when the test is used with seminary groups. The author wishes, however, to go one step further and suggest that the elimination of these items would be beneficial. If the effect of the presence of such items were merely negative, i.e., if they were merely undiagnostic and nothing more, there would be no harm in allowing them to remain. What, however, if such items are not items that are neutral but rather prejudicial to the effective operation of the test? Such, it is submitted, is the case with the seminary group because the number of unsuitable items is sufficiently large to produce an atmosphere of artificiality and unreality inimical to the test operation. (3, p. 606).

Since the publication of that suggestion, Bier has constructed an abbreviated MMPI for use with seminarians and candidates for the priesthood. The Bier modification is being used rather widely with candidates and seminarians today. Many, this writer among them, consider the use of the Bier modification to be

an unfortunate solution to the difficulties of applying the genuine MMPI to seminarians -- for the simple reason that those who use a modification cannot benefit fully from the rich and constantly increasing body of studies devoted to the full-length MMPI. A more acceptable solution to the admitted problem of applying the MMPI to seminary groups would perhaps be to construct, through empirical item-analysis, a new "seminary adjustment" scale to be added to the standard clinical scales according to which the MMPI is scored.

Among those who have followed Bier in modifying the MMPI are Benko and Nuttin, who developed "an adaptation of the test for a population of European culture, and more specially for Belgian university students" (2, p. ix). Having modified the test to suit European culture, Benko and Nuttin further modified it to make it, as they believe, more suitable for seminarians. Following Bier's suggestion, they changed the wording (and, in many cases, the content) of items especially related to religion and to social activities and attitudes. They contend that their modified version is essentially the same test as the full-length English version of the MMPI despite the fact that they have dropped 188 items and re-worded 25 others. They applied their modified version to 181 students of philosophy and theology in religious orders and to 79 novices belonging to different religious congregations of men. To these same subjects they also applied a vocational adjustment self-rating scale which was to serve as an external criterion of adjustment to seminary life. They found that their seminarians obtained significantly higher scores than their control group (soldiers and students) on only two scales, Mf and Hy (2, p. 72). They further found that three MMPI scales discriminated between well adjusted and poorly adjusted seminarians better than did the other seven clinical scales:

"The Mf scale seems of little value for our purpose. The Pt, Sc, and Hy scales, on the contrary, are those which, for our group of seminarians, are the most symptomatic. In other terms, it is tendencies toward psychasthenia, schizophrenic or schizoid personality, and hypochondriasis which are found in most characteristic fashion among seminarians maladjusted to religious life" (2, p. 101). They even single out affirmative answers to four items on the MMPI as being particularly suggestive of vocational maladjustment: "Il y a quelque chose qui cloche dans mon esprit;" "La plupart du temps j'ai le cafard;" "Je deviens anxieux et bouleverse de devoir faire un petit voyage hors de chez moi;" and "La vie est pour moi presque toujours un effort." They note that a negative answer to the statement "Je me sens heureux la plupart du temps" is similarly suggestive of vocational maladjustment (2, pp. 101-2). Finally, they suggest general norms for interpretation of seminarians' MMPI profiles: "For the group of seminarians or young religious, as for the group of novices whom we have examined, the fact of obtaining abnormally elevated results on more than two clinical scales of the MMPI seems to be a very serious indication of a lack of vocational adaptation. Elevated results on only one scale do not suffice for the elaboration of a diagnosis. They constitute an interesting indication for a more profound examination and, eventually, for appropriate direction and psychological reeducation" (2, p. 102).

CHAPTER III

DESIGN OF THE RESEARCH

Subjects. The subjects of this experiment (henceforth to be called the Experimental Group) are all members of the same large religious order of men within the Roman Catholic Church. The order is a "clerical" order (i.e., an order some of whose members receive ecclesiastical ordination). All the subjects of this study are either candidates for the priesthood or newly ordained priests completing their seminary studies. For this reason, they are referred to in the title of this thesis as "religious" seminarians (i.e., candidates for the priesthood who are members of a religious order -- as contrasted with candidates for the diocesan or "secular" clergy, who belong to no religious order and take no religious vows). All subjects are being trained in seminaries of the religious order located in the middlewestern section of the United States of America. They have all received a unified ascetical training considered by the Church to be uniquely distinctive of the religious order to which they belong. They have all been ecclesiastically approved for the traditional perpetual vows of poverty, chastity, and obedience, and all have pronounced such vows in addition to a fourth vow binding them to live perpetually according to the plan of life distinctive of the religious order of which they are members.

The religious order from which these subjects were drawn prescribes a rigidly unified course of training for its members throughout the world. That

course of training comprises two years of novitiate (devoted to intensive ascetical training); two years of juniorate (devoted to studies approximating those of the liberal arts college); three years of the study of scholastic philosophy; three years of practice teaching at the high school or college level; four years of study of dogmatic and moral theology and Canon Law; and a final year of ascetical training. Members of the order are ordained to the priesthood at the end of the third year of the study of theology. Because of the uniformity of their training, the subjects of this study were considered a priori to be a more homogeneous group of seminarians than those of the Bier study (cf. p. 8). Their homogeneity in MMPI performance required testing by a procedure to be outlined later (cf. p. 52).

These subjects were considered to be at least minimally adjusted to the life of this religious order and its seminaries for the reason that all had "persevered" (to use the term consecrated by usage in this particular order) through at least five years of training -- and some had persevered through more than 14 years. It seems not unwarranted to consider these subjects psychologically adjusted to their form of life, since it is a fact that the vast majority of men who survive five years of training in the order (which is noted for its severity in screening out inadequate members) spend the remainder of their lives in the order. It may safely be predicted that only an insignificant percentage of the seminarians who served as subjects for this study (perhaps 5 of the subjects, or approximately 7%) will drop out of the order before or after ordination to the priesthood. (This figure is based upon a study of unpublished percentages of defections from the order during the past 25 years.) On the other hand, the maladjusted or unsuitable members drop out in

comparatively great numbers during the first five (and especially during the first two) years of training. When the writer of this paper speaks, then, of his subjects as a group of "adjusted" seminarians, he means nothing more than that they are seminarians who have lived from five to more than 14 years within the order, with the approval of its demanding superiors, and that most of them may be expected (on the basis of statistical studies) to be ordained to the priesthood at the end of 13 years of training and to spend the remainder of their lives as priests within the order.

Subjects were obtained by the investigator during visits to several seminaries of the order. Superiors permitted him to address the seminarians, to explain the nature of his project, and to enlist their voluntary cooperation. No pressure was put upon any seminarian to participate in the project, either by superiors of the order or by the investigator. The seminarians were informed that the purpose of the project was the formulation of a group profile of "adjusted seminarians" of the order. It was explained to them that they qualified as "adjusted seminarians" in the minimal sense of the word explained above. They were told that only group results would be published in any form and that it was hoped that the research might serve an ultimate purpose of contributing to the construction of specific norms for the use of the MMPI in the screening of candidates for the priesthood in the religious order of which they are members. They were assured that their participation in the project would be kept completely anonymous. They were told to put no other identification on their papers except a numerical indication of their position along the continuum of years of training within the order. They were assured that the project had the complete approval of their provincial and local superiors on

condition that participation be completely voluntary.

Materials for the project were made available in sufficient quantity so that any who were even remotely interested in participating might have such materials at hand. Many more persons took such material than actually participated in the research. They were allowed to take the MMPI in the privacy of their own rooms and were urged not to discuss the test among themselves until all the testing had been completed. Those who decided ultimately not to participate merely returned blank answer sheets to a locked box placed in an unobserved location as a receptacle for completed (or blank) answer sheets. They were allowed several days to answer the test.

When the returns had been tabulated, it was discovered that 79 seminarians had filled out inventories. Of the 79 protocols, six were discarded because they showed T-scores higher than 70 on the K scale. Since Bier's study had rejected any protocols with T-scores above 70 on a validating scale, it was decided that the same procedure ought to be adopted for this research. (It is to be noted, however, that Bier did not use the K scale in his research, which was completed before the K scale was in general use. Consequently, Bier's rejections were made on the basis of excessive T-scores on scales L or F or on the basis of more than 50 unanswered items constituting the ? score. This investigator found no scores higher than 70 on scales L or F and no paper with anywhere near 50 unanswered items.) The discarding of the six protocols with high K scores left a total N of 73 for the Experimental Group. The 73 subjects of the Experimental Group were distributed, according to approximate age and precise number of years of training completed within the order, in the proportions shown in Table 1 (p. 57). Ages were estimated on the basis of known

statistics regarding mean ages for the seminarians at each point of the training; the mean age for seminarians entering the novitiate of the order is currently 19 years. The total Experimental Group was eventually divided into two subgroups for purposes of this research in accordance with a rationale to be explained later (p. 53).

Materials. Materials for the research were the booklet form of the MMPI, designed for group administration, and the IBM answer sheets which have been designed for use with the booklet form. The booklet form of the test contains 566 statements preceded by the directions:

This inventory consists of numbered statements. Read each statement and decide whether it is true as applied to you or false as applied to you.

You are to make your answers on the answer sheet you have. Look at the example of the answer sheet shown at the right. If a statement is TRUE or MOSTLY TRUE, as applied to you, blacken between the lines in the column headed T. . . . If a statement is FALSE or NOT USUALLY TRUE, as applied to you, blacken between the lines in the column headed F. . . . If a statement does not apply to you or if it is something that you don't know about, make no mark on the answer sheet.

Remember to give YOUR OWN opinion of yourself. Do not leave blank spaces if you can avoid it.

Scoring materials consisted of hand-scoring stencils for the group form of the MMPI for two validating scales (F and K) and 10 clinical scales (Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma, and Si). The other two validating scales (L and ?) are scored by inspection. All materials used are published by the Psychological Corporation.

Procedure. Each answer sheet was scored by hand for each of the four validating scales and for each of the 10 clinical scales named above. This procedure yielded raw scores on each of the 14 scales. These raw scores were

then translated into T-scores with a mean of 50 and a sigma of 10 by the use of the formula:

$$T = 50 + \frac{10 (X_1 - M)}{SD}$$

These T-scores may also be read directly from tables given in the MMPI Manual (27). Next the raw scores on the five clinical scales which are ordinarily corrected by the addition of some percentage of the K score were so corrected, and corresponding T-scores were calculated for these corrected raw scores by the use of transformation tables given in the MMPI Manual (27). The procedure thus far left the investigator with four sets of scores: raw scores without K; T-scores corresponding to raw scores without K; raw scores corrected by addition of K or some percentage thereof; T-scores corresponding to raw scores corrected for K.

As the next step in the procedure, the total Experimental Group was divided into two subgroups so that the hypothesis of homogeneity of performance on the MMPI might be tested. Inspection of Table 1 will show that the distribution of subjects according to approximate age and precise number of years of training within the order is unequal, with the greater numbers of subjects falling at the extremes. Members of the order who were at the intermediate stages of training were not readily available to the investigator in any large numbers. It was decided that the homogeneity of this group in MMPI performance might be statistically tested by dividing its members into two subgroups with a cutting-point fixed at the psychological midpoint of the course of training. This psychological midpoint was considered to be the period of practice teaching to which all prospective candidates for the priesthood in this religious order are subjected after the completion of seven of their 14 years of training. It

is generally maintained by members of the order that the practice teaching period effects more noticeable changes in the personalities of the seminarians than does any other period in the course of training. The subjects were therefore divided into those who had had at least one year of practice teaching and those who had had none. Those who had had no teaching experience will be referred to henceforth as the Non-Teaching Group; the rest will be referred to as the Teaching Group. When the total Experimental Group is so divided, N for the Non-Teaching Group becomes 36; N for the Teaching Group, 37. For the comparison of these two groups, raw scores were employed because they were readily available and directly comparable and undistorted by any transformation. The range, the mean, the standard deviation, the critical ratio (or t-ratio), and the probability of the critical ratio were calculated for these two groups on each of the 10 clinical scales. The formula used for the calculation of the critical ratio was:

$$\frac{M_1 - M_2}{\sigma_{\text{diff}}}$$

In this formula, $M_1 - M_2$ is the difference between the means of the two sub-groups (Teaching and Non-Teaching); σ_{diff} is the standard error of the difference between the means. The probability of the critical ratio was read directly from the normal probability table; such procedure was thought valid because the degrees of freedom were 34 for the Non-Teaching Group and 35 for the Teaching Group (N - 2 for each group because of the fixing of the mean and standard deviation for each group). A two-tailed test was employed because the direction of possible deviation of one group from the other was unpredictable

in advance. The .05 level of confidence was accepted in advance.

In order to test the second hypothesis -- that the Experimental Group did not differ from the Minnesota Male Normal Group -- raw scores without the K correction were employed because norms for the Minnesota Male Normal Group were available in terms of such raw scores. Means, standard deviations, and critical ratios (or t-ratios) were again calculated for each of the 10 clinical scales according to the procedure outlined immediately above. The probability of the critical ratios was again determined from the normal probability table because the N for each group was sufficiently large to warrant such procedure. A one-tailed test was used because Bier had already established that all differences between his seminarians and his college population were in the direction of higher scores for the seminarians. Since studies of college students in turn show that they deviate from the Minnesota Normal Group in the direction of higher scores (22), it was hypothesized that the Experimental Group of this thesis would, a fortiori, deviate from the Minnesota Normals in the direction of higher scores. It was decided that the .05 level of confidence would be accepted.

In testing the third and final hypothesis -- the hypothesis of no difference between the Bier Group and the Experimental Group -- T-scores were employed because the Bier results were available only in the form of T-scores calculated by the formula given on page 52. The T-scores were calculated from raw scores without the K correction because Bier did not use K in his calculations. Comparisons were made on only nine of the clinical scales because the tenth scale (Si) was not in use at the time of the Bier study. Means, standard deviations, and critical ratios (or t-ratios) were calculated for the

nine scales. The critical ratios were subjected to a two-tailed test because no hypothesis could be formulated regarding the direction of possible deviation of the Experimental Group from the Bier Group. Probability was determined from comparison with the normal probability table because the degrees of freedom were large (71 for the Experimental Group; 169 for the Bier Group). It was decided that the .05 level of confidence would be accepted as significant.

CHAPTER IV

ANALYSIS OF RESULTS

The first question which the design of the research proposed to investigate was the homogeneity of the Experimental Group. As has been stated earlier, the writer of this thesis (in order to test Bier's statement that his heterogeneous group was "a good representative sampling of students for the priesthood") chose a group which seemed prima facie more homogeneous (members of only one religious order drawn from a comparatively restricted geographical area of the United States and subjected to a uniform and highly distinctive ascetical training). It was proposed that the homogeneity of this group in MMPI performance might be statistically tested by dividing its members into two subgroups with a cutting-point fixed at the psychological midpoint of the course of training (cf. p. 52). The subjects were therefore divided into those who had had at least one year of practice-teaching and those who had not. Those who had not had any teaching experience are called the Non-Teaching Group; the rest, the Teaching Group. When the total Experimental Group is so divided, N for the Non-Teaching Group becomes 36; N for the Teaching Group, 37. When tabulated according to approximate age and years of training within the order, the subjects are grouped as shown in Table 1.

Table 1

Distribution of Experimental Group (N=73) According
to Approximate Age and Years of Training

Non-Teaching Group (N=36)

Years of Training	Age	Number of Subjects
5	24	17
6	25	16
7	26	3

Teaching Group (N=37)

Years of Training	Age	Number of Subjects
8	27	1
9	28	1
10	29	0
11	30	4
12	31	8
13	32	0
14	33	23

Inspection of Table 1 will show that the distribution is unequal, with the greater number of subjects falling at the extremes in terms of age and years of training. Members of the order who were at the intermediate point of their training were unavailable to the investigator except in small numbers. It is not felt, however, that this unequal distribution works against the testing of homogeneity; on the contrary, the investigator believes that if groups at extremes of age and training within the order do not differ significantly from each other, the total Experimental Group which they constitute may well be said to be homogeneous. Specifically, if groups of subjects at these extremes perform homogeneously on the MMPI, they may be

considered to constitute one homogeneous group for comparison with the MMPI performance with the Bier heterogeneous group.

Table 2

Differences between Means of Non-Teaching Group (N=36) and Teaching Group (N=37) in Terms of MMPI Raw Scores without K

Scale		Range	Mean	SD	t-ratio	p***
Hs	N-T*	0-13	3.69	2.82		
	T**	0-13	5.16	3.56	1.91	>.05
D	N-T	11-29	18.71	4.54		
	T	11-37	19.78	5.75	1.06	>.10
Hy	N-T	15-28	21.56	3.01		
	T	11-32	21.85	4.20	.60	>.10
Pd	N-T	10-26	16.30	3.93		
	T	6-25	15.87	4.16	.73	>.10
Mf	N-T	20-42	30.17	6.00		
	T	21-42	28.58	2.50	1.25	>.10
Pa	N-T	7-17	11.42	2.50		
	T	6-16	10.63	2.44	1.20	>.10
Pt	N-T	4-32	12.00	6.71		
	T	2-35	11.73	6.56	.34	>.10
Sc	N-T	4-35	12.33	7.08		
	T	3-27	11.00	5.62	1.18	>.10
Ma	N-T	11-27	15.95	3.02		
	T	9-24	15.46	3.70	1.08	>.10
Si	N-T	12-45	25.56	7.28		
	T	11-47	26.43	7.77	.46	>.10

*N-T indicates the Non-Teaching Group

**T indicates the Teaching Group

***is calculated for a two-tailed test

Table 2 shows that the Experimental Group does, in fact, perform homogeneously on all scales of the MMPI. It is of some interest to note that the Non-Teaching Group shows a slight (but statistically insignificant) trend toward higher scores on the "psychotic" scales of the MMPI (Pa, Sc, and Ma), while the Teaching Group shows a similar slight (but statistically insignificant) trend toward higher scores on the "neurotic" scales (Hs, D, and Hy). The only difference even approaching statistical significance is on scale 1 (Hs), on which the Teaching Group scores higher. If a one-tailed test had been used, the Teaching Group would have scored higher than the Non-Teaching Group (therefore showing greater maladjustment of a hypochondriacal nature) with a probability of .0281. Since, however, no published studies on the MMPI with which the writer of this thesis is acquainted show a trend toward elevation of Hs scores with age, there was no justification for predicting higher scores for the older group -- and, consequently, no justification for a one-tailed test.

In summary, then, it may be said that, on the basis of the data given in Table 2, the experimental group of this thesis performs homogeneously on the MMPI. Since the differences in performance between the two subgroups of the Experimental Group (Teaching and Non-Teaching) prove to be statistically insignificant, the distinction between these two subgroups will be disregarded in all further analysis of results of this research, and the Experimental Group will be treated as one whole (N=73).

Figure 1 shows the profile formed by mean raw scores uncorrected by K and transformed according to the formula described above (p. 52).

Figure 1

Profile of Experimental Group (N=73) Based upon Transformation of Raw Scores (x) Without Addition of K

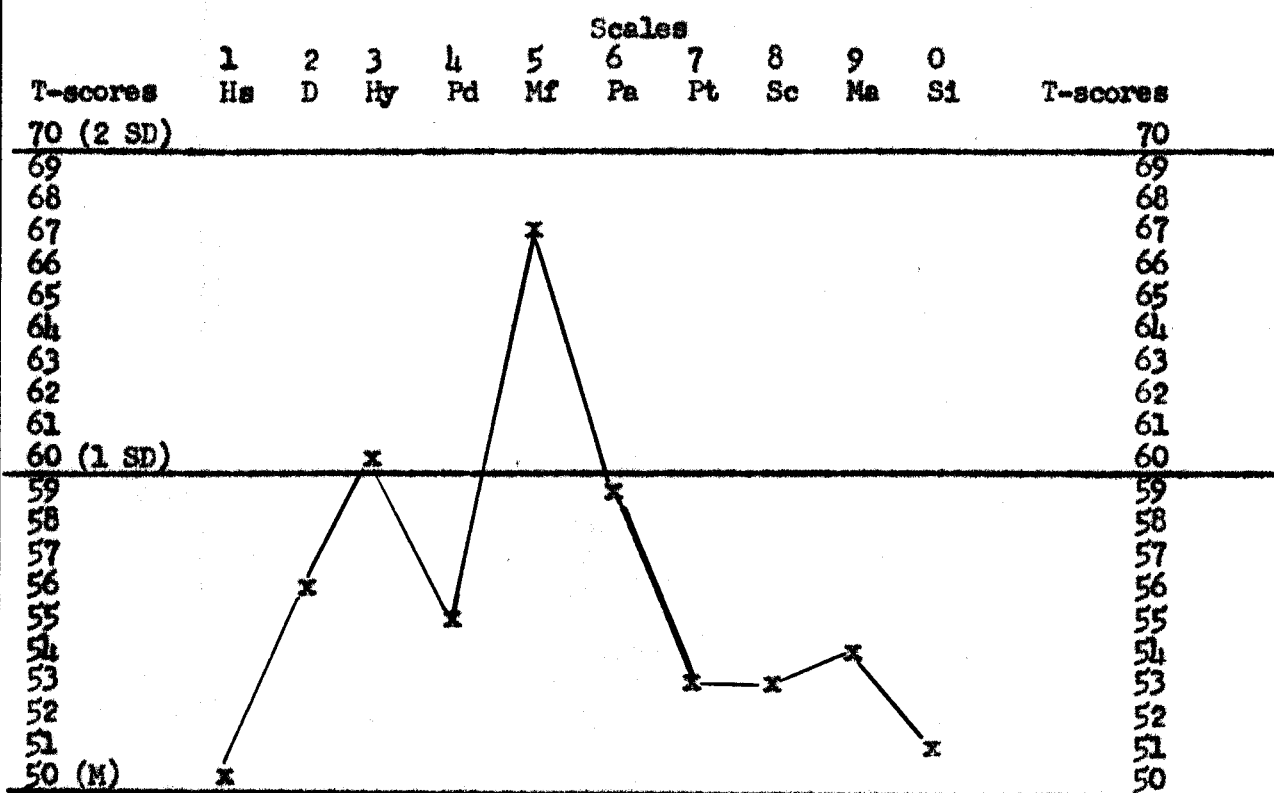
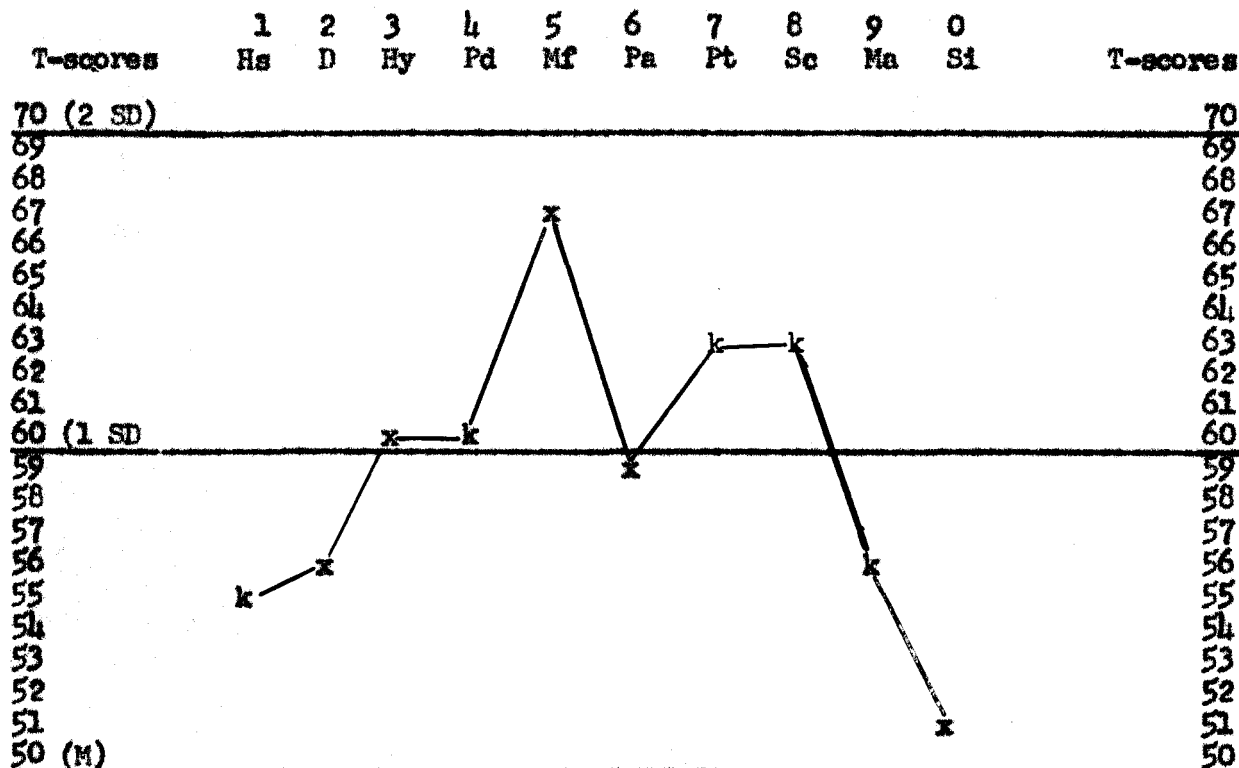


Figure 1 shows that the highest point of the group profile based on raw scores without K is equivalent to a T-score of 67 (almost 2 sigmas above the mean of Minnesota Normal Males) on scale Mf. This elevation on the Mf scale is in keeping with the findings reported in studies of male college students (22). It is now generally believed that the Mf scale measures not "femininity" so much as cultural interests more fully developed than those of the Minnesota Normal Males. Since the Mf scale is now generally discredited (cf. pp. 29-30), it may be more meaningful to consider the second-highest elevation shown in Table 1 as the most distinctive MMPI score of the

Experimental Group. The second-highest scale is the Hy scale, which is elevated to precisely 1 standard deviation above the mean for Minnesota Normal Males. This group of seminarians would, therefore, appear to be markedly more tense than the Minnesota Normal Males, with their tensions seeking outlet in conversion symptoms such as gastrointestinal complaints, cardiac manifestations, paralyses, contractures, etc. (27, p. 19). The next highest elevation in the group profile occurs on scale 6 (Pa), on which the group mean falls almost 1 standard deviation above the mean for Minnesota Normal Males. The MMPI Manual describes subjects with high scores on scale 6 as "characterized by suspiciousness, oversensitivity and delusions of persecution" (27, p. 20) when the scores are in the abnormal range (i.e., higher than a T-score of 70). Since the Experimental Group score on the Pa scale is not within the abnormal range, Hathaway's description of the milder symptoms found among the Pa criterion group may be more useful as a clue to the meaning of the Pa performance of these seminarians. "Milder symptoms," says Hathaway, "included suspiciousness, an excess of interpersonal sensitivity, and an underlying rigidity of opinions and attitudes" (26, p. 110). The Experimental Group also appears to be characterized by a higher than usual degree of depression (D scale) and by an unusual absence of deep emotional response (Pd scale). As calculated on the basis of raw scores without the K correction (Table 3), the profile does not seem to be significant in any other respects (since T-scores between 54 and 46 are generally disregarded in the interpretation of profiles). When coded by the original Hathaway system (25; 27, p. 17), the group profile based upon transformed raw scores without K would be written as '(5)3624--.

Figure 2

Profile of Experimental Group (N=73) Based upon Transformation of Raw Scores With Addition of K (k)



When the K correction is taken into account (as in Figure 2), the group profile changes quite radically. The Mf scale remains predominant, but scales 7 (Pt) and 8 (Sc) are elevated into a tied second place. This would suggest that, if the K correction is valid for this group, the group is characterized (on the basis of the high Pt score) by "phobias or compulsive behavior" and by "mild depression, excessive worry, lack of confidence, or inability to concentrate" (27, p. 20). On the basis of an equally high Sc score, the group may be said to be "characterized by bizarre and unusual thought or behavior"

(27, p. 20) of a type commonly thought of as "schizoid." If these interpretations of the group characteristics of these seminarians are valid (and their validity hinges upon the validity of the K correction), they are indeed disturbing. For the K correction makes the Pt, Sc, and Hy scales the high points of the profile. But these are precisely the scales which Benko and Nuttin have found to be symptomatic of maladjustment to religious life (cf. p. 46)! It has been proved, however, that the Pt scale correlates very highly with Sc, so that elevation on the one is apt to elevate the other concomitantly (cf. p. 35). Moreover, T-scores below 70 are considered within the normal range, so that no point of the profile of the Experimental Group (even when modified by K) may be considered "dangerously" high. Still, elevation on the three scales which Benko and Nuttin have found to be diagnostic of maladjustment to seminary life suggest that either Benko and Nuttin's modified MMPI is not equivalent to the full-length MMPI used in this study (which is quite possible); or that the Experimental Group is not an adjusted group (and that, consequently, a group of seminarians may go on living in a religious order for years and perhaps for a lifetime and be basically maladjusted to their way of life -- which seems absurd); or, finally, that the K correction distorts the MMPI performance of this seminary group.

There is some question in the investigator's mind about the propriety of using the K correction with this Experimental Group. Since the individual subjects took the MMPI with assurance of complete anonymity, with rather full knowledge of the purposes which their protocols were to serve, and with no pressure of any sort exerted upon them to participate in the project, it seems quite unlikely that any of them should reasonably have felt any compulsion to

"fake good" (as the consecrated MMPI terminology expresses the tendency to avoid looking abnormal on the test). Nevertheless, their mean score on the K scale was 17 (for both Teaching and Non-Teaching Groups). A score of 17 lies approximately 1 sigma above the mean for Minnesota Normal Males on the K scale. McKinley and Hathaway report: "The largest mean that we have observed on the K scale was obtained from graduate electrical engineers. These men were studied during the war and were mostly around 30 years of age. They were exempted from military duty in order to carry on aviation research and at the time of testing were applying for special airplane control testing at high altitudes" (37, p. 120). These men had obvious reasons for "faking good." Since the mean for the Experimental Group of this thesis is even higher than this "largest mean" (16.72), the Experimental Group seems clearly unusual in its K performance. For that reason, and since the writer of this thesis can offer no explanation for the anomaly except to hazard the guess that age and education (and particularly test sophistication) may have raised the K scores, he prefers to leave the interpretation of the K-corrected scores to possible later research and to consider as the more valid profile -- or at least as the more certainly intelligible profile -- that which has been drawn without K (Figure 1).

Benko and Nuttin report that "abnormally elevated results on more than two clinical scales of the MMPI seems to be a very serious indication of a lack of vocational adaptation" (2, p. 102; cf. p. 46). Table 3 will show the number of elevated profiles for the Experimental Group.

Table 3

Number of Protocols of Experimental Group (N=73) with Abnormal Elevations ($T > 70$) on Basis of Scores with K-Correction Added

No scale elevated	33
Only Mf elevated	11
1 scale elevated	10
2 scales elevated	5
3 scales elevated (incl. Mf)	7
4 scales elevated	2
5 scales elevated	1
6 scales elevated	2
7 scales elevated	1
8 scales elevated	1

Since all subjects of the Experimental Group who showed elevation on three scales showed elevation on the Mf scale (which Benko and Nuttin admit to be an insignificant scale -- cf. p. 46), only seven subjects may be considered to show elevation on "more than two" scales. Seven subjects constitute approximately 10% of the Experimental Group. If the Benko and Nuttin findings are valid for their modified MMPI, and if their modified MMPI is equivalent to the original MMPI used in this study, 10% of the Experimental Group are apt to be badly adjusted to seminary life. On the basis of statistical expectation (cf. p. 48), this figure seems high by about 3%.

The next step in the analysis of data involves making a comparison between the Experimental Group and the Minnesota Normal Male Group. Table 4 shows such a comparison. The computation of Table 4 was complicated by the unfortunate fact that standardization data for some of the scales of the MMPI are available only in somewhat defective form. The N of the standardization

Table 4

Comparison of Experimental Group (N=73) with Minnesota Normal Males
on Basis of Raw Scores without K

Scale	N	Mean	SD	t-ratio	p***
Hs M*	300 (est.)****	4.5	4.5		
E**	73	4.45	3.19	.110	>.10
D M	347	18.1	4.9		
E	73	19.25	5.15	1.748	<.05
Hy M	345	16.50	5.50		
E	73	21.70	3.60	10.097	<.01
Pd M	294	12.99	4.00		
E	73	16.08	4.04	7.571	<.01
Mf M	300 (est.)	20.50	5.00		
E	73	29.38	5.33	12.926	<.01
Pa M	300 (est.)	8.00	3.5		
E	73	11.03	2.47	8.584	<.01
Pt M	293	10.00	7.00		
E	73	11.87	6.64	2.130	<.05
Sc M	300 (est.)	9.50	7.50		
E	73	11.67	6.35	2.523	<.01
Ma M	294	14.51	4.42		
E	73	15.71	3.36	2.553	<.01
Si M	300 (est.)	25.00	9.00		
E	73	26.00	7.53	.978	>.05

*M indicates the Minnesota Normal Male Group

**E indicates the Experimental Group of this thesis

***p was calculated for a one-tailed test

****est. indicates an estimated N for the Minnesota Normal Male Group

sample of Minnesota Normal Males has not been published for all scales. Since those Ns which are precisely known are 293, 294, 294, 345, and 347 -- and since

it is known that approximately the same number of males were used for the standardization of each scale -- the investigator felt that he would not be doing violence to the data by estimating an N of 300 for those scales for which the precise N was unknown. A comparison of the Experimental Group and the Minnesota Normal Male Group seemed desirable, if not imperative, and no other way of making such a comparison was available. However, conclusions drawn from Table 4 should be read with the reservation that they are only as precise as the estimate of N for those scales for which N was estimated. (If N were actually as low as 200 for the scales for which it was estimated at 300, only the differences on the Pt scale would drop from an acceptable to an unacceptable degree of confidence.)

On the assumption that the estimates in Table 4 are substantially accurate, it appears that -- even at the .01 level of confidence -- the Experimental Group differs significantly from the Minnesota Male Normal Group on scales 3 (Hy), 4 (Pd), 5 (Mf), 6 (Pa), 8 (Sc), and 9 (Ma). The differences on scales 2 (D) and 7 (Pt) are significant beyond the .05 level of confidence. Only on scales 1 (Hs) and 10 (Si) are the differences insignificant even at the .05 level. The norms by which the MMPI is usually scored are, then, obviously inapplicable to this seminary population except in the cases of the scales for hypochondriasis and social introversion.

Finally, the comparison may be made between the MMPI performance of this Experimental Group and the MMPI performance of the Bier Group of seminarians. It has been demonstrated that the Experimental Group is homogeneous in its performance on the MMPI and that it differs from the Minnesota Normal Male Group performance very significantly (at the .01 level of confidence) on six

of the 10 scales and significantly (at the .05 level of confidence) on two additional scales. Now the final question to be answered is: whether this Experimental Group of seminarians from a religious order represents the same population as the Bier seminary sample, which has been called "a good representative sampling of students for the priesthood." Table 5 presents the data on the basis of which an answer to that question may be made.

In the calculation of probabilities in Table 5, a two-tailed test was employed, since it was impossible to predict the direction in which the Experimental Group might differ from the Bier Group if it differed at all. It is evident, however, from Table 5 that in all instances of significant difference, the difference is in the direction of higher scores for the Experimental Group.

If the .05 level of significance is accepted, the Experimental Group performs in a significantly different manner from the Bier Group on four scales: scales 3 (Hy), 4 (Pd), 5 (Mf), and 6 (Pa). Of these four significantly different performances, two -- the performances on scales 5 (Mf) and 6 (Pa) are significant even at the .01 level of confidence. Greatest discrepancy in performance is found on scale 6 (Pa).

How are these data to be interpreted? If Hathaway's description of the type of personality that scores moderately highly on scale 6 (Pa) is correct, then the Experimental Group is marked by "suspiciousness, an excess of interpersonal sensitivity, and an underlying rigidity of opinions and attitudes" (26, p. 110) -- and this to a significantly greater degree than the Bier group. In accordance with the common current interpretation of the significance of the Mf scale, it may also be said that the Experimental Group is significantly

Table 5

Comparison of Experimental Group (N=73) and Bier Group (N=171)
on Basis of T-Scores Transformed from Raw
Scores without K

Scale		Mean	SD	t-ratio	p***
Hs	E*	50.12	7.76	.882	>.10
	B**	51.09	8.14		
D	E	55.94	12.58	1.462	>.10
	B	53.42	11.74		
Hy	E	59.16	6.64	2.139	<.05
	B	57.01	7.72		
Pd	E	55.42	10.43	2.180	<.05
	B	52.30	9.78		
Mf	E	67.41	10.38	2.802	<.01
	B	63.44	9.52		
Pa	E	59.75	6.75	5.611	<.01
	B	54.33	7.27		
Pt	E	53.18	9.19	1.074	>.10
	B	53.92	8.95		
Sc	E	53.37	3.05	.570	>.10
	B	53.10	7.70		
Ma	E	53.87	7.70	.712	>.10
	B	54.66	8.44		

*E indicates Experimental Group of this thesis

**B indicates Bier Seminary Group

***p was calculated for a two-tailed test

more cultured in its tastes and interests than is the Bier Group. If the .05 level of confidence is accepted, the higher performance of the Experimental Group on scale 3 (Hy) may be considered to indicate that members of the

Experimental Group are more marked than the Bier Group by "somatic complaints," by a tendency to consider themselves "unusually well socialized," by periods of unhappiness and "blues" (35, p. 90). On the basis of the high score on scale 4 (Pd), the experimental group may be considered significantly less inhibited, less conforming, and emotionally shallower than the Bier Group (27, p. 19).

Interpretation of scores within the normal range -- between T-scores of 50 and 70 -- in which all the Experimental Group means lie is an extremely delicate and hazardous operation. The most that can be said with certainty is that the performance of the Experimental Group is significantly different (in a statistical sense) from that of the Bier Group at the .01 level of confidence on scales which are designed to identify paranoid and effeminate personalities and at the .05 level of confidence on scales which are designed to identify hysteric and psychopathic personalities. But these scales clearly identify such personalities, if at all, only in cases where the T-scores exceed 70. Moreover, even with T-scores above 70, all scales are fallible in the sense that they may misclassify a certain percentage of normals who produce "false positive" records: 6% for the Pd scale, 3% for the Hy scale, and an unspecified percent for the Pa scale. The Mf scale has no known clinical significance, as this paper has repeatedly insisted. Consequently, the only prudent statement that can be made about the data of Table 5 in a qualitative sense is that the Experimental Group of this thesis performs in ways somewhat similar to the ways in which paranoid, effeminate (or highly cultured), hysteric, and asocial psychopathic personalities perform on the MMPI -- to a degree significantly greater than the degree to which the Bier Group performs in these ways.

The major null hypothesis of this thesis was the hypothesis of no difference in MMPI performance between the Bier Group and the Experimental Group significant at the .05 level of confidence. On the basis of the data presented above, that null hypothesis may be rejected -- specifically for scales 3, 4, 5, and 6. In other words, the MMPI performance is so significantly different that the two groups may be said to be samples of significantly different populations. Consequently, Bier's contention that his subjects represent "a good representative sampling of students for the priesthood" must be accepted with qualification. The statement cannot be read to mean that the Bier Group may be used as a normative group for all seminary populations; it certainly may not be used as normative for scales 3, 4, 5, and 6. The subjects who participated in this study were all young men who had lived from five to 14 years within the one religious order of which they all were members. They had all been ecclesiastically approved for perpetual vows within that order. Some of them (N=23) had already been ordained to the priesthood. A study of defections from the order within the past 25 years leads to the statistical expectation that approximately 93% of the subjects of this experiment will live out their entire lives as priests in this religious order. These facts suggest that this group may be considered -- as a group -- well adjusted to seminary and priestly and religious life. Yet on four scales of the MMPI, this group performs in a statistically significant "abnormal" way. The subjects appear statistically to be significantly "maladjusted" when compared even with another seminary group. It cannot be said from this study how early such "deviations" begin to appear in members of this religious order; a study of younger members than those employed in this research would help to

answer that question. Nor can any statement about the degree of adjustment to religious or seminary or priestly life be made about any individual member of this group; a study using some external criterion of vocational adjustment would help to answer that question. But the results of this study do suggest that any religious order (and perhaps any seminary) which contemplates using the MMPI as a screening device for candidates for the priesthood should construct its own norms, using adjusted members of its own specific population as a standardization group or normative sample, rather than simply accepting either the Bier results or the results of the study of the Experimental Group of this thesis, since there appears to be no such thing (at least to date) as the identifiable "seminarian profile" for the MMPI.

CHAPTER V

SUMMARY

The investigator proposed to study the MMPI performances of a homogeneous group of seminarians, members of one distinctive religious order within the Roman Catholic Church. He proposed to compare their MMPI performance with the performances of the Minnesota Male Normal Group (used as the standardization sample for the MMPI) and of the Bier heterogeneous group of seminarians. He administered the MMPI to 79 anonymous volunteers who were presumably well adjusted -- as a group -- to seminary life within their religious order. He rejected the protocols of six of these seminarians because their T-scores on the K scale exceeded 70. N therefore equalled 73.

He discovered no significant intra-group differences in MMPI performance within his Experimental Group.

He found significant differences between the performance of his Experimental Group and the performance of the Minnesota Male Normal Group at the .01 level of confidence on scales 3, 4, 5, 6, 8, and 9; at the .05 level of confidence on scales 2 and 7. Only on scales 1 and 0 were the performances statistically undistinguishable.

He found significant differences between the performance of his Experimental Group and the performance of the Bier Group of seminarians at the .01 level of confidence on scales 5 and 6; at the .05 level of confidence on scales 3 and 4.

His Experimental Group yielded distinctly different MMPI profiles depending upon whether those profiles were drawn from scores with or without the K correlation added. Without the K correction, the high points of the group profile were scales 5, 3, 6, 2, and 4 (in that order). With the K correction, the high points were scales 5, 7 & 8 (tied), 3 & 4 (tied), 6, 2 & 9 (in that order).

The investigator advanced the suggestion that his results indicate the need for religious orders and seminaries to construct their own individual norms if they intend to use the MMPI as a screening device for candidates, since his study indicates that there is no one identifiable "seminarian profile" for the MMPI.

REFERENCES

1. Acta et documenta congressus generalis de statibus perfectionis. Rome, 1950.
2. Benko, A., and Nuttin, J. Examen de la personnalité chez les candidats à la prêtrise, adaptation du test de personnalité MMPI. Louvain: Publications Universitaires, 1956.
3. Bier, W. C. A comparative study of five Catholic college groups on the MMPI. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis, Univer. of Minnesota Press, 1956. Pp. 589-609.
4. Bier, W. C. Psychological testing of candidates and the theology of vocation. Rev. for Religious, 1953, 12, 291-304.
5. Bier, W. C. Practical requirements of a program for the psychological screening of candidates. Rev. for Religious, 1954, 13, 13-27.
6. Biot, R. Psychonevroses et séminaires. Bull. des Anciens Elèves de St. Sulpice, 1935.
7. Biot, R., and Galimard, P. Guide médical des vocations sacerdotales et religieuses. Paris: Editions Spes, 1945.
8. Burke, H. R. Personality traits of successful minor seminarians. Washington: Catholic Univer. Press, 1947.
9. Cahen-Salabelle, R. Vocations et affectivité. Cahiers Laënnec, 1950, 10, 3-30.
10. Codex iuris canonici.
11. Cossa, P. Fausses vocations et vocations troublées. Rev. de Droit Canonique, 151, 1, 63-68.
12. Couly, A. Inaptitudes, étude canonique des textes anciens et du Codex. Le Recrutement Sacerdotal, 1927, nouvelle série 1, 5-24.
13. De la Croix, M. E. La vocation sacerdotale. Paris: Maison de Bon Pasteur, 1926.
14. Digna, M. (Sister). A tentative testing program for religious life. Rev. for Religious, 1951, 10, 75-76.

15. Drake, L. E. Scale 0 (social introversion). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 181-183.
16. Drake, L. E., and Thiede, W. B. Further validation of scale 0 (Si). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 184-186.
17. Duffey, F. D. Testing the spirit. St. Louis: Herder, 1950.
18. Eck, D., and Larere, C. Psychasténie et vocation. Cahiers Laennec, 1950, 10, 3-17.
19. Elaine, M. (Mother). Influence of religion on personality patterns. ACPA Newsletter Supplement, 1957, No. 29.
20. Ernst, P. Option vitale: contribution à une psychologie ascétique de la vocation. Nouvelle Rev. Théologique, 1947, 79, 731-742, 1065-1084.
21. Geraud, J. Contre-indications médicales à l'orientation vers le clergé. Lyons-Paris, 1947.
22. Goodstein, L. D. Regional differences in MMPI responses among male college students. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 574-578.
23. Gough, H. G. Some common misconceptions about neuroticism. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 51-57.
24. Guadier, G. Les irrégularités "ex defectu corporis." Paris, 1933.
25. Hathaway, S. R. A coding system for MMPI profile classification. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 126-129.
26. Hathaway, S. R. Scales 5 (masculinity-femininity), 6 (paranoia), and 8 (schizophrenia). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 104-111.
27. Hathaway, S. R., and McKinley, J. C. Manual for the MMPI. New York: Psychological Corporation, 1951.

28. Hathaway, S. R., and McKinley, J. C. Construction of the schedule. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 60-63.
29. Hathaway, S. R., and McKinley, J. C. Scale 2 (depression). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 73-80.
30. Hathaway, S. R., and Meehl, P. E. An atlas for the clinical use of the MMPI. Minneapolis: Univer. of Minnesota Press, 1951.
31. Hathaway, S. R., and Monachesi, E. D. (Eds.). Analyzing and predicting juvenile delinquency with the MMPI. Minneapolis: Univer. of Minnesota Press, 1953.
32. Johnson, Ruth L. E. The relation of religious attitudes and selected personality characteristics. M.A. thesis, Univer. of Minnesota, 1948.
33. McCarthy, T. J. Personality traits of seminarians. Studies in Psychol. and Psychiat. from the Catholic Univer. of Amer., 1942, No. 4.
34. McKinley, J. C., and Hathaway, S. R. Scale 1 (hypochondriasis). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 64-72.
35. McKinley, J. C., and Hathaway, S. R. Scales 3 (hysteria), 9 (hypomania), and 4 (psychopathic deviate). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 87-103.
36. McKinley, J. C., and Hathaway, S. R. Scale 7 (psychasthenia). In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 81-86.
37. McKinley, J. C., Hathaway, S. R., and Meehl, P. E. The K scale. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 112-123.
38. Meehl, P. E. The dynamics of "structured" personality tests. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 5-11.

39. Meehl, P. E., and Hathaway, S. R. The K factor as a suppressor variable in the MMPI. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 12-40.
40. Moore, T. V. Insanity in priests and religious. Ecclesiastical Rev., 1936, 95, 485-498, 601-613.
41. Mulders, A. La vocation au sacerdoce. Bruges, 1925.
42. Nabais, J. Contribution de la psychologie et de la pedagogie à la selection et l'orientation des vocations sacerdotales. Unpublished licentiate paper, Univer. of Louvain, 1953.
43. Pius XI, Pope. Christian education of youth (Divini illius magistri). New York: Paulist Press, no date.
44. Pius XI, Pope. The Catholic priesthood (Ad catholici sacerdotii). New York: America Press, 1936.
45. Pius XII, Pope. Holiness in priestly life (Menti nostrae). New York: Paulist Press, 1951.
46. Seeman, W. Subtlety in structured personality tests. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956, Pp. 41-50.
47. Sempe, L. Vocation. In Dictionnaire de la théologie Catholique, XV, col. 3153.
48. Simoneaux, H. J. Spiritual guidance and the varieties of character. New York: Pageant Press, 1956.
49. Sinety, R. Psychopathies: sur leur prodrômes chez les candidats au sacerdoce. Le Recrutement Sacerdotal, 1927, nouvelle serie 1, 25-34.
50. Welsh, G. S., and Dahlstrom, W. G. (Eds.). Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956.
51. Wheeler, W. M., Little, K. B., and Lehner, G. F. The internal structure of the MMPI. In G. S. Welsh and W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 255-263.

APPROVAL SHEET

The thesis submitted by Patrick John Rice, S.J. has been read and approved by three members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

June 1958
Date

Frank J. Hobler
Signature of Adviser